Adaptive Estimation for Financial Time Series

D I S S E R T A T I O N

zur Erlangung des akademischen Grades
doctor rerum politicarum
(dr. rer. pol.)
im Fach Wirtschaftswissenschaften

eingereicht an der
Wirtschaftswissenschaftliche Fakultät
Humboldt-Universität zu Berlin

von
Herrn Dipl.-Vw. Danilo Mercurio
geboren am 06.03.1975 in Venaria Reale (I)

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. J. Mlynek

Dekan der Wirtschaftswissenschaftliche Fakultät:
Prof. Dr. J. Schwalbach

Gutachter:

1. Prof. Dr. W. Härdle
2. Prof. Dr. V. Spokoiny

eingereicht am:	10. Mai 2004
Tag der mündlichen Prüfung:	11. Juni 2004

Abstract

This thesis develops new *locally adaptive* methods for estimation and forecasting of financial time series data. These methods are mainly tailored for volatility estimation of financial returns and for regression and autoregression problems. The proposed approaches are defined locally adaptive because instead of imposing a stationary data generating process which can be globally described by a finite number of parameters, they only assume that observations which are chronologically close to each other can be well approximated by a constant process. These procedures are adaptive in the sense that for each observation they choose in a data driven way the *interval of time homogeneity*, i.e. the number of chronologically close and homogeneous past data where the hypothesis of a constant structure can not be rejected. Nonasymptotic theoretical results are derived, which show the optimality of the suggested algorithms. Comparisons with standard approaches demonstrate that the new procedures behave competitively and offer a valuable alternative, furthermore, intensive simulation studies and applications to real data provide good results, confirming their effectiveness and practical relevance.

Keywords:
adaptive estimation, local homogeneity, financial data, forecasting

Zusammenfassung

Diese Dissertation entwickelt neue *lokal adaptive* Methoden zur Schätzung und Vorhersage von Zeitreihendaten. Diese Methoden sind für die Volatilitätsschätzung von Finanzmarktrenditen und für Regressions- und Autoregressionsprobleme konstruiert worden. Die vorgeschlagenen Ansätze werden als lokal adaptiv bezeichnet, denn, anstatt einen globalen datenerzeugenden Prozess aufzuzwingen, welcher durch eine endliche Anzahl von Parametern beschrieben werden kann, nehmen sie nur an, daß Beobachtungen, welche chronologisch nah bei einander liegen, durch einen konstanten Prozess gut approximiert werden können. Diese Prozeduren sind adaptiv, weil sie für jede Beobachtung in einer datengesteuerten Art und Weise das *Intervall der Zeithomogenität*, d.h. die Anzahl der chronologisch benachbarten und homogen vergangenen Daten, aussuchen, für welchen die Hypothese einer konstanten Struktur nicht verworfen werden kann. Nichtasymptotische theoretische Ergebnisse werden hergeleitet, welche die Optimalität der betrachteten Algorithmen zeigen. Vergleiche mit Standardansätzen verdeutlichen, daß die neuen Prozeduren sich kompetitiv verhalten und eine nützliche Alternative bieten, außerdem liefern intensive Simulationsstudien und Anwendungen an reellen Daten gute Ergebnisse und bezeugen dabei ihre Effektivität und praktische Relevanz.

Schlagwörter:
adaptive Schätzung, lokale Homogenität, Finanzmarktdaten, Vorhersage

Acknowledgement

This dissertation has considerably benefited from the suggestions and comments of many of colleagues and friends which I would like to thank. First of all, I would like to express my deep gratitude to my supervisors who have the merit of teaching me how to do research and who have transmitted to me their enthusiasm for this activity. Prof. Vladimir Spokoiny constant support and encouragement was extremely precious and the joint work with him was fundamental for the realization of this book. Moreover, the scientific cooperation with prof. Wolfgang Härdle, his advises and his remarks have proven to be very useful and stimulating.

The working atmosphere at the Institute of Statistics and Econometrics at the Department of Economics of the Humboldt University, at the Wierstraß Institute and at the Center for Applied Statistics and Economics have positively contributed to the development of this thesis. For their comment and suggestions I would like to thank: Y. Chen, K. Detlefsen, R. Ihle, M. Krätzig and M. Reiß. Furthermore, a special thank is deserved by my coauthor prof. C. Torricelli.

Finally, I must not forget to mention my parents, whose help, support and guidance were extremely valuable and significantly contributed to the realization of my work.


C.A.S.E.
CENTER FOR APPLIED STATISTICS AND ECONOMICS

Contents

List of Figures

List of Tables

List of Frequent Notation

$X \equiv Y$	X has the constant value Y
$X \stackrel{\text{def}}{=} Y$	X is defined as Y
$X \sim F$	the random variable X has distribution F
$X \approx Y$	X is approximately equal to Y
$\mathsf{P}(A)$	probability of the set A
$\mathsf{E}(X)$	expected value of the random variable X
$\mathcal{F}_t$	sigma algebra
$\mathsf{E}_t(X)$ or $\mathsf{E}(X\|\mathcal{F}_t)$	conditional expectation of the random variable X
$\mathsf{Var}(X)$	variance of the random variable X
S_t	stock price or exchange rate
$R_t = \ln(S_t/S_{t-1})$	log-return
σ_t	standard deviation
Σ_t	covariance matrix
$[a,\ b]$	closed interval
$[a,\ b[$	right half open interval
$[x]$	integer part of the scalar x
$\|x\|$	absolute value of the scalar x
I or J	time interval
$\|I\|$ or N_I	number of points in I
$\mathcal{I}$	set of time intervals
$\mathcal{J}(I)$	set of time intervals J, such that $J \subset I$
$\mathbb{I}$	optimal interval of homogeneity
$\mathcal{T}_I$	set of testing point
$T_{\tau,I}$ and T_I	test statistics

ν_k	stopping time
$\mathrm{N}(\mu,\ \sigma^2)$	normal distribution with mean μ and variance σ^2
$\mathsf{tr}\Sigma$	trace of the matrix Σ
$\mathsf{det}\Sigma$	determinant of the matrix Σ
$\mathsf{diag}X$	diagonal matrix with the vector X on the main diagonal
$X^\top$	transpose of X
$L_I(\theta)$	log likelihood functions for the observations in the time interval J
$\widehat{L}_I$	log likelihood evaluated at its maximum
$K(\theta, \theta')$	Kullback-Leibler information for the two normal distributions with variances θ and θ'
ACF	autocorrelation function
(G)ARCH	(Generalized) Autoregressive Conditional Heteroskedasticity
LAVE	Local Adaptive Volatility Estimate
LCPD	Local Change Point Detection
LR	Likelihood Ratio
MLE	Maximum Likelihood Estimator
OLS	Ordinary Least Square
VaR	Value at Risk

Chapter 1

Introduction

1.1 Motivation

The aim of this thesis is to provide a contribution to the statistical analysis of financial market data. A remarkable amount of statistical research is devoted to financial time series. The interest in this topic is motivated by the needs of the financial industry, which uses estimated parameters in order to practically implement theoretical models of pricing, hedging, trading and risk management. Moreover, a sound statistical evaluation of the data is encouraged and even imposed by the regulators, which require financial institutions to develop and implement probabilistic models for the assessment of market risk, credit risk and operational risk. Errors in estimation or in the choice of the appropriate statistical model can be very costly, resulting in unhedged positions, misperception of risks and increased capital requirement.

This research focuses on the issue of volatility estimation and on its implication for risk management, furthermore the question of portfolio hedging is analyzed. In order to tackle these problems, new nonparametric locally adaptive methods are developed. The proposed estimators are called locally adaptive because instead of imposing a stationary data generating process which can be globally described by a finite number of parameters, they only assume that observations which are chronologically close to each other can be well approximated by a constant process.

In order to emphasize the novelty of the approach, it is appropriate to confront it with the standard way of modeling economic and financial time series. We therefore provide a brief summary of the main ideas of parametric time series analysis, while keeping the discussion to an informal and intuitive level since a precise and complete comparison lays beyond the scope of this dissertation. Further details and references to this topic can be found for example in Hamilton (1994).

1.1.1 The classical approach

At least since the work of Box and Jenkins (1976) traditional time series econometrics strongly relies on the assumption of (covariance) stationarity. This assumption is very useful because it leads to a parsimonious parameterization of the time series and it implies that the properties of the estimators can be studied via a well known asymptotic statistical theory. Stationarity means that the observed time series is interpreted as the realization of a stochastic process Y_t, whose mean and autocovariance do not depend on the date t, i.e.:

$$\begin{aligned} \mathsf{E}(Y_t) &= \mu \qquad \text{for all } t \\ \mathsf{E}(Y_t - \mu)(Y_{t-j} - \mu) &= \gamma_j \qquad \text{for all } t \text{ and any } j, \end{aligned}$$

in particular this means that the variance γ_0, must be constant over time. This assumption ensures that a time series Y_t can be meaningfully modeled by an autoregressive moving average process of the form:

$$Y_t = \nu + \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \xi_t + \theta_1 \xi_{t-1} + \ldots + \theta_q \xi_{t-q},$$

where ξ_t is a white noise innovation process, such that $\mathsf{E}(\xi_t) = 0$ and $\mathsf{Var}(\xi_t) = \sigma^2$ for all t.

Conversely, the parameterization described by the equation above leads to a stationary process if the roots of the characteristic polynomial:

$$1 - \phi_1 y - \phi_2 y^2 - \ldots - \phi_p y^p$$

lay outside the unit circle. If instead some roots are on the unit circle the process may exhibit a trend behavior, like for example gross domestic product time series, or a variance which is increasing over time, like for example exchange rate time series. In this case the series can be made stationary by differencing.

Many financial time series though, such as interest rates, inflation rates, exchange rate returns and stock returns clearly display a non-constant variance (Figure 1.1) and this phenomenon, also known as volatility clustering, apparently contradicts the hypothesis of stationarity; see Franke et al. (2003) for a review of the stylized facts of financial data. In order to reconcile this fact within the stationarity framework one can model the conditional variance σ_t of the innovation ξ_t by a generalized autoregressive conditional heteroskedasticity process (GARCH):

$$\mathsf{E}_{t-1}(\xi_t) = \sigma_t \quad \text{and} \quad \sigma_t^2 = \omega + \alpha_1 \xi_{t-1}^2 + \ldots + \alpha_p \xi_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_q \sigma_{t-q}^2.$$

Given certain technical conditions on the parameters, the above process is stationary and it implies an autoregressive moving average representation for the square innovations ξ_t^2. GARCH models were first proposed by Bollerslev (1986) as a generalization of the ARCH models of Engle (1982) and they soon became very popular because, allowing for a constant variance and a time varying conditional variance, they provide a simple parametric explanation of the volatility clustering phenomenon, which can be handled with the usual tools. A large number of papers has followed the first publications on this topic, and the original models have been extended in order to provide better explanations of the features of the data. For example, models which take into account asymmetries in volatility have been proposed, such as EGARCH (Nelson, 1995), QGARCH (Sentana, 1995) and GJR (Glosten et al., 1992); furthermore, the research on integrated processes has produced integrated (Engle and Bollerslev, 1986) and fractal integrated versions of the GARCH model.

In general, stationary parametric processes can be estimated for example by the maximum likelihood method or the generalized method of moments,

and asymptotic properties of the estimators can be derived. If the sample size of the data is sufficiently large, the limiting distributions can be used in order to make inferential statements about the estimated parameters. The availability of very large samples of financial data, often at daily frequency, has made it possible to construct models which display rich and quite complicated parameterizations. However, the proliferation of many different, concurring models may be also interpreted as a sign that a finite number of parameters does not provide a good description of the data, which are possibly characterized by structural instability. For example from Figure 1.1 one can see that fitting a GARCH model to a long data set and to two subsamples of the same data produces results which are optically quite different. A detailed analysis of these questions can be found in Mikosch and Starica (2000a).

The large econometric literature on structural breaks hints to the fact that the stationarity assumption can not be taken for granted. Economies evolve and are subject to sudden shifts precipitated by legislative changes, economic policy, major discoveries and political turmoil. Occasionally, the timing of the structural break is clearly recognizable, such as in the case of the German reunification, but in the most general setting the researcher has to face a multiple change point detection problem with unknown location and number of change points.

Some of the most important references for estimation with structural breaks can be found in Lütkepohl (1988), Hamilton (1989), Hamilton (1993), Stock (1993) and Clements and Hendry (1999). Although the list above is not complete, it is somehow representative of the evolution of the econometric way of thinking about structural instability. Indeed, the older literature tends to present the parameter shifts as occasional events, while the more recent approaches model the parameters themselves as stochastic processes (usually random walks or finite state Markov chains) and suggest intercept correction and overdifferencing methods for improving the forecasting in the case structural breaks.

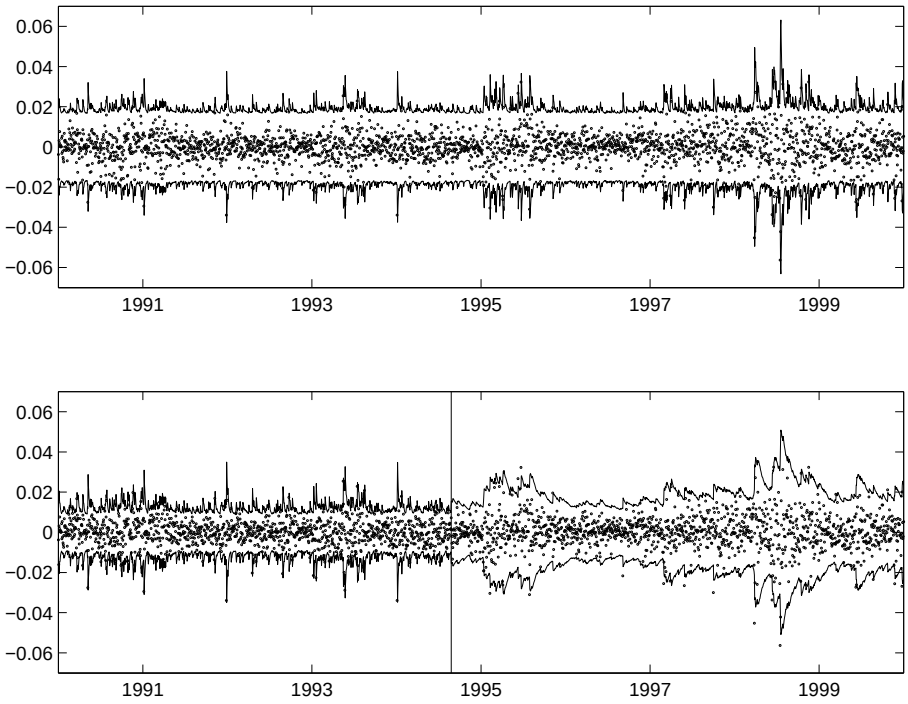

Figure 1.1: USD/JPY exchange rate (dots) and conditional standard deviation estimated with a GARCH(1,1) (solid line). The upper plot shows a global estimate, while in the lower plot the estimation was performed separatly over the two subsamples.

1.1.2 The nonparametric adaptive approach

An alternative to the global parametric modeling is offered by the nonparametric statistics. The main idea of this approach consists in modeling only the local structure of the data. The hypothesis that the model can be completely described by a finite number of parameters is dropped and it is assumed that only subsamples of observations which are "close" to each other follow a homogeneous model.

Many techniques developed in the nonparametric statistics community were originally concerned with the problems of density estimation and nonlinear regression, see for example Härdle et al. (2004), and subsequently some of those approaches have been generalized in order to handle time series data.

One of the estimation procedures from classical nonparametric statistics which is closest in spirit to the ones presented in this thesis is kernel regression. Instead of assuming a finite global parametric model, kernel regression specifies some general features of the regression function, such as the degree of smoothness. The function is then estimated locally, for example via a polynomial approximation. One of the main issues in kernel regression is the choice of the bandwidth, i.e. the smoothing parameter which defines the neighborhood of the data where the local approximation is good. The smoothness assumptions on the regression function (for example the existence of bounded first derivatives) can be imposed on the whole data set, i.e. globally, or only on subsets of the data, i.e. locally. In this latter case, the regression function may be characterized by a finite number of singularities and/or jumps at unknown locations. This singularity points represent natural borders for the neighborhood of data where an approximation with a smooth function, like a polynomial, is appropriate and therefore their detection has to be kept into account in the estimation algorithm.

This problem is analyzed among others by Lepski (1990), Lepski and Spokoiny (1997) and Spokoiny (1998), who develop adaptive pointwise methods for the estimation of a regression function with nonhomogeneous smoothing characteristics. Their approach, which is analogous to the one presented

in the following chapters, is pointwise adaptive in the sense that the regression function is estimated at each point independently and the estimator adapts automatically to the local degree of smoothness. Alternative approaches based on wavelet regression are discussed for example by Donoho et al. (1994) and by Wang (1995).

1.1.3 Adaptive estimation for time series data

In a time series context the data are naturally ordered in a chronological way and the assumption of parametric statistics may be relaxed by imposing a model which is only locally stationary, in the sense that only observations which are close to each other chronologically are considered homogeneous. This approach leads to estimators which are based on time varying coefficient modeling. In the remainder of this section we provide a brief overview of the literature on this topic which is mostly related to the dissertation.

Fan and Zhang (1999), Cai et al. (2000a) and Cai et al. (2000b) assume that the model parameters smoothly vary over time and can be locally approximated by a linear function of time. Fan et al. (2001) discuss applications of this method to continuous time diffusion models for stock prices and interest rates and they propose kernel regression techniques for the estimation of the time varying coefficient functions in the drift and in the volatility. Dahlhaus and Rao (2003) apply the locally stationary approach to the discrete time GARCH model and suggest a time varying coefficient estimator. Similarly, Fan and Gu (2003) compare different adaptive volatility estimators and evaluate their performance using both statistical tests and criteria based on the Value at Risk. Cheng et al. (2003) focus on the nonparametric filtering problem and on its applications to volatility estimation. They consider the choice of the filtering parameters, which can be made globally or locally, where the local version means that at each time point t the filter parameter is chosen in order to optimize the performance near t, while the global version selects the filtering parameters on a "large" presample and then uses the selected parameters for forecasting the postsample values. This situation is analogous to the local bandwidth and global bandwidth selection

in the nonparametric smoothing literature, see for example Brockmann et al. (1993) and Fan and Gijbels (1995).

As previously reported, the GARCH model is one of the most popular global parametric models for the volatility of financial time series, nevertheless many recent papers challenge its ability to provide sensible estimation result for long data samples. For example, Mikosch and Starica (2000b) assert that structural breaks are a major source of problems for GARCH models. Starica (2003), comparing the forecasting performance of the GARCH model with the one of a nonparametric regression approach, finds out that the two methods are equivalent for the short horizon and that the GARCH is outperformed for the long horizon. This lack of forecasting ability for long horizons is interpreted by the author as a sign of nonstationarity.

The studies above and several others like for example Lamureux and Lastrapes (1990), Simonato (1992) and Kleibergen and Dijk (1993) document that parameter instability, structural breaks and nonstationarity can be expected when using the GARCH model for financial time series. Therefore the development of accurate change point detection algorithms appears to be a crucial task in this context.

The change point detection problem for financial time series is considered in Mikosch and Starica (2000a) who focus on asymptotic properties of the test if only one change point is present. Kitagawa (1987) applies non Gaussian random walk modeling with heavy tails as the prior for the piecewise constant mean for one step ahead prediction of nonstationary time series and Hamilton and Baldev (2002) present a survey of Markov Switching models for applications in economics and finance. However, the mentioned modeling approaches require some essential amount of prior information about the frequency of change points and their size.

1.2 Outline of the dissertation

This thesis studies the problem of estimation and forecasting in the presence of structural breaks and time varying parameters. Throughout the dissertation, the assumption is made that the unknown parameters can be locally approximated by a constant and procedures are suggested for the identification of the interval of time homogeneity, i.e. the subsample of chronologically close observations, where this approximation is appropriate. In particular, the focus is set on applications of financial time series and volatility models and the suggested methods are tailored for forecasting in an on-line framework. This means that, as a new observation becomes available, the model is checked and the estimate is updated. The procedures are also pointwise adaptive in the sense that at each date the estimation is performed independently, i.e. the interval of homogeneity (which for prediction purposes includes only past values) is computed and on this interval the estimate is calculated by averaging. The suggested estimation algorithms work both in

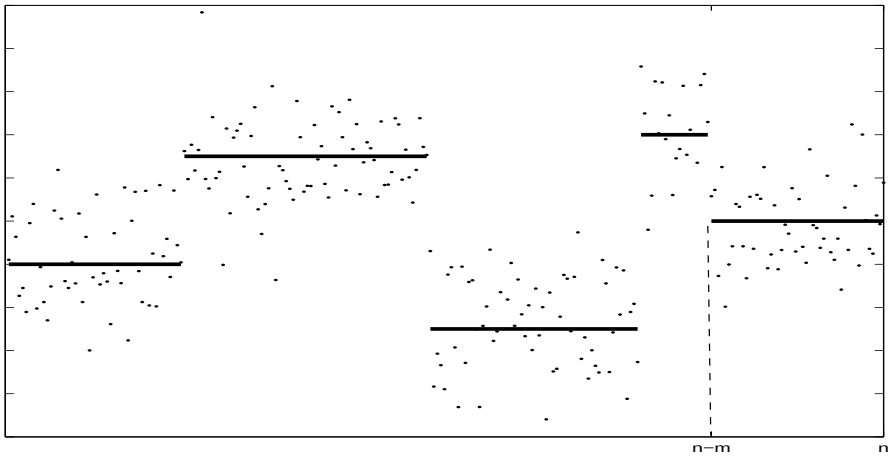

Figure 1.2: Example of a locally homogeneous process (change point model), the dots represent the observed data while the line is the target of the estimation.

the case of varying parameters which are smooth functions of time and for change point models like the one shown in Figure 1.2.

The estimation of a change point model is studied in depth both from the theoretical point of view and through intense simulation studies. For this model, the interval of time homogeneity is intuitively defined as the period between the most recent observation and the latest change point (for example $[n-m,\, n]$ in Figure 1.2) and its length is estimated via multiple testing algorithms which are roughly defined as follows:

- starting from a small time interval, whose right hand point is the date of the most recent observation, we test the null hypothesis of homogeneity against a change point alternative,
- if the hypothesis is rejected, then a change point is detected and the interval of homogeneity is estimated,
- otherwise the length of the interval candidate is increased by including less recent dates and the test is performed again,
- the procedure is repeated until a change point is detected, or until the whole data set has been tested and not rejected.

In Chapter 2 and 3 the test statistic is obtained from an exponential probability bound for martingales by Lipster and Spokoiny (1999), while in Chapter 4 the test is derived via the likelihood ratio approach.

Non-asymptotic Bonferroni methods are used in order to study the theoretical properties of the multiple testing algorithms. The theoretical results are obtained imposing only very mild conditions on the unknown parameters, so that the change point model and the model where the parameters are smooth functions of time are treated in a unified way. The estimators are shown to be optimal in the sense that the quality of the adaptive estimate is proved to be of the same order of the one obtained with the prior knowledge of the interval of time homogeneity. Furthermore, for the special case of the change point model it is shown that the suggested procedures deliver a nearly optimal quality of change point detection as presented by Brodskij and Darkhovskij (1993).

The critical values of the testing procedures which represent the most important tuning parameters are obtained by simulation in order to ensure that the wrong rejection of an homogeneous interval, i.e. the type-1-error, is kept under control. Finally, applications to real data in the context of volatility estimation, portfolio selection and Value at Risk confirm the practical relevance of the proposed algorithms.

This thesis is made up of three self contained essays, which have been written in collaboration with V. Spokoiny and C. Torricelli and is organized as follows:

Chapter 2: Statistical inference for time inhomogeneous volatility models. A new approach for estimation and forecasting of the volatility of financial time series is proposed. No assumption is made about the parametric form of the processes, on the contrary we only suppose that the volatility can be approximated by a constant over some interval. In such a framework the main problem consists in filtering this *interval of time homogeneity*. Afterwards, the estimate of the volatility can be simply obtained by local averaging. We construct a *locally adaptive volatility estimate (LAVE)* which can perform this task and investigate it both from the theoretical point of view and through Monte Carlo simulations. Finally the LAVE procedure is applied to a data set of nine exchange rates and a comparison with a standard GARCH model is also provided. Both models appear to be able of explaining many of the features of the data, nevertheless the new approach seems to be superior to the GARCH method as far as the out of sample results are taken into consideration.

Chapter 3: Estimation and arbitrage opportunities for exchange rate baskets. This chapter analyzes short term portfolio investment opportunities in a capital market where a currency is defined as a currency basket. In line with the mean-variance hedging approach, we determine a self financed optimal investment strategy which minimizes the expected quadratic cost function. The successful implementation of the speculative strategy requires a precise estimate of the basket weights which are possibly

non-constant over time. To this end, we suggest an adaptive nonparametric procedure which provides satisfactory results both on simulated and real data. We apply the optimal investment strategy to the case of the Thai Baht basket whereby the weights are computed by means of the adaptive estimator. We also implement a recursive estimator, a rolling estimator and the Kalman filter, which serve as benchmark models and compare our results with the literature. The different estimators are evaluated with profit based criteria and the performance of the adaptive estimator turns out to be the best one.

Chapter 4: Estimation of time dependent volatility via local change point analysis. In this chapter, the same testing problem as for Chapter 3 is analyzed. However, here a multiple testing procedure based on the likelihood ratio test is developed, which is suitable, both for univariate and multivariate volatility estimation. The procedure recovers this interval from the data using the local change point analysis. Afterwards the estimate of the volatility can be simply obtained by local averaging. An application to exchange rate data, and a comparison with the LAVE procedure and with a standard GARCH model are provided. The numerical results demonstrate a reasonable performance of the new method, which is suitable for risk management applications. Indeed, one can obtain the Value at Risk, for each series and for a portfolio, from the empirical distribution of the standardized residuals. A backtesting exercise shows that the values computed with this method are conservative but precise, in accordance with the requirements of the regulators.

1.3 Concluding remarks

In this dissertation new nonparametric *locally adaptive* methods for estimation and forecasting of financial time series data are developed. These methods focus on volatility estimation of financial returns and regression and autoregression problems. No assumption is made about the parametric structure of the processes of interest, which are estimated locally via a

constant approximation. This approach allows one to treat in a unified way the estimation of smoothly varying functions of time and of change point models. Moreover, the choice of a simple local model makes it possible to construct a procedure which automatically selects the subsample of the data where the approximation with a constant function is correct. The estimation is performed sequentially, only with past data, and therefore it is suitable for forecasting purposes. The nonparametric approaches are compared with standard parametric methods: GARCH(1,1) model for the volatility and the Kalman filter and they behave competitively providing a valuable alternative. In particular, the volatility estimator is able to remove the variance clustering effect from the data, so that the distribution of the devolatilized residuals, and the VaR, can be successfully estimated as if the data were independent and identically distributed. Moreover, the local constant volatility estimation, unlike the GARCH model, can be applied to multivariate models of moderate dimension like the nine dimensional exchange rate data set considered in this research.

Chapter 2

Statistical inference for time-inhomogeneous volatility models

2.1 Introduction

A remarkable amount of statistical research is devoted to financial time series, in particular, to the volatility of asset returns, where the term volatility indicates a measure of dispersion, usually the variance or the standard deviation. The interest in this topic is motivated by the needs of the financial industry, which regards volatility as one of the main reference numbers for risk management and derivative pricing.

Actually, asset returns time series display very peculiar stylized facts, which are connected with their second moments. Graphically, they look like white noise, where periods of high and low volatility seem to alternate. Their density has fat tails if compared to that of a normal random variable, and they show significantly positive and highly persistent autocorrelation of the absolute returns, meaning that large (resp. small) absolute returns are likely to be followed by large (resp. small) absolute returns. Typical examples can be seen in Section 2.6, and further details on this topic can be found in Taylor (1986). Therefore, a white noise process with time varying variance

is usually taken to model such features. Let S_t denote the observed asset process, then the corresponding (log) returns $R_t = \ln(S_t/S_{t-1})$ follow the heteroskedastic model

$$R_t = \sigma_t \xi_t,$$

where ξ_t are standard Gaussian independent innovations and σ_t is a time varying *volatility* coefficient. It is often assumed that σ_t is measurable w.r.t. the σ-field generated by the preceding returns $R_1, \ldots, R_{t-1}$. For modeling this volatility process, parametric assumptions are usually used. The main model classes are the ARCH and GARCH family (Engle, 1995), and the stochastic volatility (Harvey et al., 1995). A large number of papers has followed the first publications on this topic, and the original models have been extended in order to provide better explanations. For example, models which take into account asymmetries in volatility have been proposed, such as EGARCH (Nelson, 1995), QGARCH (Sentana, 1995) and GJR (Glosten et al., 1992); furthermore, the research on integrated processes has produced integrated (Engle and Bollerslev, 1986) and fractal integrated versions of the GARCH model.

The availability of very large samples of financial data has made it possible to construct models which display quite complicated parameterizations in order to explain all the observed stylized facts. Obviously those models rely on the assumption that the parametric structure of the process remains constant through the whole sample. However, if this assumption is not fulfilled, the resulting estimates are biased and the forecasting performance can be quite poor (Clements and Hendry, 1998). Furthermore, checking for parameter instability becomes quite difficult if the model is nonlinear and/or the number of parameters is large. Thus, those characteristics of the returns, which are often explained by the long memory and (fractal) integrated nature of the volatility process, could also depend on the parameters being time varying.

Here, another approach is proposed which focuses on a very simple model but with a possibility for model parameters to depend on time. This means

that the model is regularly checked and adapted to the data. No assumption is made about the parametric structure of the volatility process, it is only supposed that it can be locally approximated by a constant, that is, for every time point n there exists a past interval $[n-m, n]$ where the volatility σ_t did not vary much. This interval is referred to as *interval of time homogeneity*. An algorithm is proposed for data driven estimation of the interval of time homogeneity, after which the estimate of the volatility can be simply obtained by averaging.

This approach is similar to varying coefficient modeling from Fan and Zhang (1999), see also Cai et al. (2000a) and Cai et al. (2000b). Fan et al. (2001) discuss applications of this method to stock price volatility modeling. They propose a procedure which is based on the assumption that the model parameters smoothly vary with time and can be locally approximated by a linear function of time and therefore it has drawback of not allowing one to incorporate structural breaks in to the model.

Change point modeling with applications to financial time series is considered in Mikosch and Starica (2000a). Kitagawa (1987) applies non Gaussian random walk modeling with heavy tails as the prior for the piecewise constant mean for one step ahead prediction of nonstationary time series. However, the mentioned approaches require some essential amount of prior information about the frequency of change points and their size.

The LAVE approach does not assume smooth or piecewise constant structure of the underlying process and does not require any prior information. The procedure proposed below in Section 2.3 focuses on adaptive choice of the interval of homogeneity that allows to proceed in a unified way with smoothly varying coefficient models and change point models.

The proposed approach attempts to describe the *local* dynamic of the volatility process, and it is particularly appealing for short term forecasting purposes which is an important building block, for example in Value at Risk and portfolio hedging problems or backtesting (Härdle and Stahl, 2000).

The remainder of the chapter is organized as follows. The next section introduces the adaptive modeling procedure. Then some theoretical properties are discussed in the general situation and for a change point model. A simulation study illustrates the performances of the new methodology with respect to the change point model. The question of selecting the smoothing parameters is also addressed and some solutions are proposed. Finally, the procedure is applied to a set of nine exchange rates and it appears to be highly competitive with standard GARCH(1,1), which is used as a benchmark model.

2.2 Modeling volatility via power transformation

Let S_t be an observed asset process in discrete time, $t = 1, 2, \ldots, n$ and R_t are the corresponding returns: $R_t = \ln(S_t/S_{t-1})$. We model this process via the *conditional heteroskedasticity* assumption

$$R_t = \sigma_t \xi_t \,, \tag{2.1}$$

where ξ_t, $t \geq 1$, is a sequence of independent standard Gaussian random variables and σ_t is the *volatility* process which is in general a predictable random process, that is, σ_t is measurable with respect to $\mathcal{F}_{t-1}$ with $\mathcal{F}_{t-1} = \sigma(R_1, \ldots, R_{t-1})$ (the σ-field generated by the first $t-1$ observations).

A *time homogeneous* (*time homoskedastic*) model means that σ_t is a constant. The process S_t is then a geometric Brownian motion observed at discrete time moments. The assumption of time homogeneity is too restrictive in practical applications, and it does not allow one to fit real data very well. Therefore, an approach based on the *local time homogeneity* is considered, this means that for every time moment n there exists a time interval $[n-m, n]$ where the volatility process σ_t is nearly constant. Under such a modeling, the main intention is both to describe the interval of homogeneity

and to estimate the corresponding value σ_n which can then be used for one step forecasting and the like.

2.2.1 Data transformation

The model equation (2.1) links the target volatility process σ_t with the observations R_t via the multiplicative errors ξ_t. The classical well developed regression approach relies on the assumption of additive errors which can be then smoothed out by some kind of averaging. A natural and widespread method of transforming the equation (2.1) into a regression like equation is to apply the log function to both its sides squared:

$$\ln R_t^2 = \ln \sigma_t^2 + \ln \xi_t^2, \tag{2.2}$$

which can be rewritten in the form

$$\ln R_t^2 = \ln \sigma_t^2 + C + v\zeta_t,$$

with $C = \mathsf{E} \ln \xi_t^2$, $v^2 = \mathsf{Var} \ln \xi_t^2$ and $\zeta_t = v^{-1} (\ln \xi_t^2 - C)$; see for example Gouriéroux (1997). This is a usual regression equation with the "response" $Y_t = \ln R_t^2$, target regression function $f(t) = \ln \sigma_t^2 + C$ and homogeneous "noise" $v\zeta_t$.

The main problem with this approach is due to the distribution of the errors ζ_t, which is highly skewed and gives very high weights to the small values of the errors ξ_t. In particular, this leads to a serious problem with missing data which are typically modeled equal to previous values providing $R_t = 0$.

Another possibility is based on power transformation (Carroll and Ruppert, 1988) which also leads to a regression with additive noise and this noise is much closer to a Gaussian one. Due to (2.1), the random variable R_t is conditionally on $\mathcal{F}_{t-1}$ Gaussian and

$$\mathsf{E} \left(R_t^2 \,\middle|\, \mathcal{F}_{t-1} \right) = \sigma_t^2.$$

Similarly, for every $\gamma > 0$,

$$\mathsf{E}\left(\left|R_t\right|^\gamma \,\middle|\, \mathcal{F}_{t-1}\right) = \sigma_t^\gamma \mathsf{E}\left(|\xi|^\gamma \,\middle|\, \mathcal{F}_{t-1}\right) = C_\gamma \sigma_t^\gamma,$$
$$\mathsf{E}\left(\left|R_t\right|^\gamma - C_\gamma \sigma_t^\gamma \,\middle|\, \mathcal{F}_{t-1}\right)^2 = \sigma_t^{2\gamma} \mathsf{E}\left(|\xi|^\gamma - C_\gamma\right)^2 = \sigma_t^{2\gamma} D_\gamma^2,$$

where ξ denotes a standard Gaussian r.v., $C_\gamma = \mathsf{E}|\xi|^\gamma$ and $D_\gamma^2 = \mathsf{Var}|\xi|^\gamma$. Therefore, the process $|R_t|^\gamma$ allows for the representation

$$|R_t|^\gamma = C_\gamma \sigma_t^\gamma + D_\gamma \sigma_t^\gamma \zeta_t, \tag{2.3}$$

where ζ_t is equal to $\left(|\xi|^\gamma - C_\gamma\right)/D_\gamma$. Note that the problem of estimating σ_t is in some sense equivalent to the problem of estimating $\theta_t = C_\gamma \sigma_t^\gamma$, which is the conditional mean of the transformed process $|R_t|^\gamma$. This is already a kind of heteroskedastic regression problem with additive errors $D_\gamma \sigma_t^\gamma \zeta_t$ satisfying

$$\begin{aligned} \mathsf{E}\left(D_\gamma \sigma_t^\gamma \zeta_t \mid \mathcal{F}_{t-1}\right) &= 0, \\ \mathsf{E}\left(D_\gamma^2 \sigma_t^{2\gamma} \zeta_t^2 \mid \mathcal{F}_{t-1}\right) &= D_\gamma^2 \sigma_t^{2\gamma}. \end{aligned}$$

A natural choice of the parameter γ is $\gamma = 2$ providing the nearly efficient variance estimation under homogeneity. For $\gamma = 2$ one has $C_\gamma = 1$ and $D_\gamma^2 = 2$. Note however that the distribution of the "errors" $\zeta_t = (|\xi_t|^\gamma - C_\gamma)/D_\gamma$ is still heavy tailed and highly skewed, which results in a low sensitivity of the method in an inhomogeneous situation. The other important cases are $\gamma = 1$ and $\gamma = 1/2$. A minimization of skewness $\mathsf{E}\zeta_\gamma^3$ and kurtosis $\mathsf{E}\zeta_\gamma^4 - 3$ with respect to γ leads to the choice $\gamma \approx 1/2$. The corresponding density $p_{1/2}(x)$ of $\zeta_{1/2}$ together with the conveniently scaled normal density $\phi(x)$ is plotted in Figure 2.1. Our numerical results are also in favor of the choice $\gamma = 1/2$; see Section 2.6 .

2.3 Estimation under local time homogeneity

Here, we describe one approach to volatility modeling based on the assumption of local time homogeneity starting from the preliminary heuristic discussion. The assumption of local time homogeneity means that the function

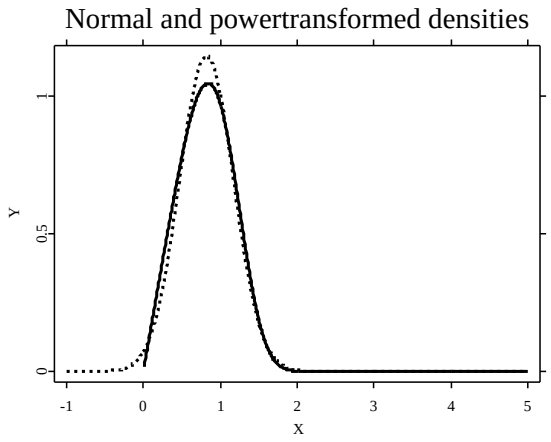

Figure 2.1: Density of $p_{1/2}(x)$ (straight line) and the normal density (dotted line).

σ_t is nearly constant within an interval $I = [n-m, n]$, and the process R_t follows the regression like equation (2.3) with the constant trend $\theta_I = C_\gamma \sigma_I^\gamma$ which can be estimated by averaging over this interval I:

$$\tilde{\theta}_I = \frac{1}{|I|} \sum_{t \in I} |R_t|^\gamma. \tag{2.4}$$

For a particular case $\gamma = 2$, the estimate $\tilde{\theta}_I$ coincides with the local maximum likelihood estimator (MLE) of the volatility σ_t^2 considered in Fan et al. (2001). As discussed in the previous section, a smaller value of γ might be preferred for improving the stability of the method. Similarly to Fan et al. (2001) one can also incorporate the one-sided kernel weighting to this estimator.

By (2.3)

$$\tilde{\theta}_I = \frac{C_\gamma}{|I|} \sum_{t \in I} \sigma_t^\gamma + \frac{D_\gamma}{|I|} \sum_{t \in I} \sigma_t^\gamma \zeta_t = \frac{1}{|I|} \sum_{t \in I} \theta_t + \frac{s_\gamma}{|I|} \sum_{t \in I} \theta_t \zeta_t \tag{2.5}$$

with $s_\gamma = D_\gamma / C_\gamma$ so that

$$\mathsf{E}\tilde{\theta}_I = \mathsf{E}\frac{1}{|I|}\sum_{t\in I}\theta_t\,, \tag{2.6}$$

$$\frac{s_\gamma^2}{|I|^2}\mathsf{E}\left(\sum_{t\in I}\theta_t\zeta_t\right)^2 = \frac{s_\gamma^2}{|I|^2}\mathsf{E}\sum_{t\in I}\theta_t^2. \tag{2.7}$$

2.3.1 Some properties of the estimate $\tilde{\theta}_I$

Due to our assumption of local homogeneity, the process θ_t is close to θ_n for all $t \in I$. Define also

$$\Delta_I = \sup_{t\in I}|\theta_t - \theta_n| \qquad \text{and} \qquad v_I^2 = \frac{s_\gamma^2}{|I|^2}\sum_{t\in I}\theta_t^2.$$

The value of Δ_I measures the departure from homogeneity within the interval I, and it can be regarded as an upper bound of the "bias" of the estimate $\tilde{\theta}_I$. The value of v_I^2, because of (2.7), will be referred as the "conditional variance" of the estimate $\tilde{\theta}_I$. The next theorem provides a probability bound for the estimation error, i.e. the deviation of $\tilde{\theta}_I$ from the present value of the volatility θ_n in terms of Δ_I and v_I.

Theorem 2.1 *Let the volatility coefficient σ_t satisfy the condition*

$$b \le \sigma_t^\gamma \le bB, \tag{2.8}$$

with some positive constant b, B. Then there exists $a_\gamma > 0$ such that for every $\lambda \ge 1$

$$\mathsf{P}\left(|\tilde{\theta}_I - \theta_n| > \Delta_I + \lambda v_I\right) \le 4\sqrt{e}a_\gamma^{-1}\lambda(1+\ln B)e^{-\lambda^2/(2a_\gamma)}.$$

Remark 2.1 *This result can be slightly refined for the special case when the volatility process σ_t for $t \in I$ is deterministic or (conditionally) independent*

of the observations R_t *preceding* I*. Namely, in such a situation, the factor* $4\sqrt{e}a_\gamma^{-1}\lambda(1+\ln B)$ *in the bound can be replaced by 2:*

$$\mathsf{P}\left(|\tilde\theta_I - \theta_n| > \Delta_I + \lambda v_I\right) \le 2e^{-\lambda^2/(2a_\gamma)}.$$

A similar remark applies to all the results what follow.

The result of this theorem bounds the loss of the estimate $\tilde\theta_I$ via the value Δ_I and the conditional standard deviation v_I. Under homogeneity, $\Delta_I \equiv 0$ and the error of estimation is of order v_I. Unfortunately, v_I depends, in turn, on the target process θ_t. One would be interested in another bound which does not involve the unknown function θ_t. Namely, using (2.7) and assuming Δ_I small, one may replace the conditional standard deviation v_I by its estimate

$$\tilde v_I = s_\gamma \tilde\theta_I |I|^{-1/2}. \tag{2.9}$$

Theorem 2.2 *Let* $R_1, \ldots, R_n$ *obey (2.1) and let (2.8) hold true. Then, for the estimate* $\tilde\theta_I$ *of* θ_n *for every* $D \ge 0$ *and* $\lambda \ge 1$

$$\mathsf{P}\left(|\tilde\theta_I - \theta_n| > \lambda' \tilde v_I,\ \Delta_I / v_I \le D\right) \le 4\sqrt{e}\lambda(1+\ln B)e^{-\lambda^2/(2a_\gamma)},$$

where λ' *solves*

$$\lambda + D = \lambda'/(1 + \lambda' s_\gamma |I|^{-1/2}).$$

2.3.2 Adaptive choice of the interval of homogeneity

Given observations $R_1, \ldots, R_n$ following the time inhomogeneous model (2.1), we aim to estimate the current value of the parameter θ_n using the estimate $\tilde\theta_I$ with a properly selected time interval I of the form $[n-m, n]$ to minimize the corresponding estimation error. Below we discuss one approach which goes back to the idea of pointwise adaptive estimation; see Lepski (1990), Lepski and Spokoiny (1997) and Spokoiny (1998). The idea of

the method can be explained as follows. Suppose I is an interval candidate; that is, we expect time homogeneity in I and hence, in every subinterval of I. This particularly implies that the value Δ_I is small and similarly for all Δ_J, $J \subset I$, and that the mean values of the θ_t over I and over J nearly coincide. Our adaptive procedure roughly means the choice of the largest possible interval I such that the hypothesis that the value θ_t is a constant within I is not rejected. For testing this hypothesis, we consider the family of subintervals of I of the form $J = [n - m', n]$ with $m' < m$ and for every such subinterval J compare two different estimates: one is based on the observations from J, and the other one is calculated from the complement $I \setminus J = [n - m, n - m'[$. Theorems 2.1 and 2.2 can be used to bound the difference $\tilde{\theta}_J - \tilde{\theta}_{I\setminus J}$ under homogeneity within I. Indeed, the conditional variance of $\tilde{\theta}_{I\setminus J} - \tilde{\theta}_J$ is $v^2_{I\setminus J} + v^2_J$ and can be estimated by $\tilde{v}^2_{I\setminus J} + \tilde{v}^2_J$. Thus, with a high probability, it holds that

$$|\tilde{\theta}_{I\setminus J} - \tilde{\theta}_J| \le \lambda \sqrt{\tilde{v}^2_{I\setminus J} + \tilde{v}^2_J},$$

provided that λ is sufficiently large. Therefore, if there exists a testing interval $J \subset I$ such that the quantity $|\,\tilde{\theta}_{I\setminus J} - \tilde{\theta}_J\,|$ is significantly positive, then the hypothesis of homogeneity is rejected for the interval I. Finally, our adaptive estimate corresponds to the largest interval I such that the hypothesis of homogeneity is not rejected for I itself and all smaller considered intervals.

Now a formal description is presented. Suppose a family $\mathcal{I}$ of interval candidates I is fixed. Each of them is of the form $I = [n-m, n]$, $m \in \mathbb{N}$, so that the set $\mathcal{I}$ is ordered due to m. With every such interval, one associates the estimate $\tilde{\theta}_I$ of θ_n and the corresponding estimate $\tilde{v}_I$ of the conditional standard deviations v_I. Next, for every interval I from $\mathcal{I}$, a set $\mathcal{J}(I)$ of testing subintervals J is considered (one example of these sets $\mathcal{I}$ and $\mathcal{J}(I)$ is given in Section 2.6). For every $J \in \mathcal{J}(I)$, we construct the corresponding estimate $\tilde{\theta}_J$ (resp. $\tilde{\theta}_{I\setminus J}$) from the observations $Y_t = |R_t|^\gamma$ for $t \in J$ (resp. for $t \in I \setminus J$) according to (2.4) and compute $\tilde{v}_J$ (resp. $\tilde{v}_{I\setminus J}$).

Now, with a constant λ, define the adaptive choice of the interval of homogeneity by the following iterative procedure:

Initialization Select the smallest interval in $\mathcal{I}$.

Iteration Select the next interval I in $\mathcal{I}$ and calculate the corresponding estimate $\tilde{\theta}_I$ and the estimated conditional standard deviation $\tilde{v}_I$.

Testing homogeneity Reject I, if there exists one $J \in \mathcal{J}(I)$ such that

$$\left|\tilde{\theta}_{I\setminus J} - \tilde{\theta}_J\right| > \lambda\sqrt{\tilde{v}^2_{I\setminus J} + \tilde{v}^2_J}. \tag{2.10}$$

Loop If I is not rejected, then continue with the iteration step by choosing a larger interval. Otherwise, set $\hat{I} =$ "the latest non rejected I".

The *locally adaptive volatility estimate (LAVE)* $\hat{\theta}_n$ of θ_n is defined by applying this selected interval $\hat{I}$:

$$\hat{\theta}_n = \tilde{\theta}_{\hat{I}}\ .$$

The next section discusses the theoretical properties of the LAVE algorithm in a general framework, while Section 2.6 gives a concrete example for the choice of the sets $\mathcal{I}$, $\mathcal{J}(I)$ and the parameter λ. This choice is then applied to simulated and real data.

2.4 Theoretical properties

In this section, we collect some results describing the quality of the proposed adaptive procedure.

2.4.1 Accuracy of the adaptive estimate

Let $\hat{I}$ be the interval selected by our adaptive procedure. We aim to show that our adaptive choice is up to some constant factor in the losses as good as the "ideal" choice $I\!I$ that may utilize the knowledge of the volatility process σ_t. This "ideal" choice can be defined by balancing the accuracy of approximating the underlying process θ_t (which is controlled by Δ_I) and the stochastic error controlled by the stochastic standard deviation v_I. By

definition, $v_I = s_\gamma |I|^{-1} \left(\sum_{t \in I} \theta_t^2 \right)^{1/2}$ so that v_I typically decreases when $|I|$ increases. For simplicity of notation, we shall suppose further that $v_I \le v_J$ for $J \subset I$.

We do not give a formal definition of an "ideal" choice of the interval I since there is no one universally optimal choice even if the process θ_t is known. Instead we consider a family of all "good" intervals $\mathit{I\!I}$ such that the variability of the process θ_t inside $\mathit{I\!I}$ is not too large compared to the conditional stochastic deviation $v_{\mathit{I\!I}}$. This, due to Theorem 2.1, allows us to bound with a high probability the losses of the "ideal" estimate $\tilde\theta_{\mathit{I\!I}}$ by $(D+\lambda)v_{\mathit{I\!I}}$ provided that $\Delta_{\mathit{I\!I}}/v_{\mathit{I\!I}} \le D$ and λ is sufficiently large. A similar property should be fulfilled for all smaller intervals $I \subset \mathit{I\!I}$. Hence, it is natural to quantify the quality of the interval $\mathit{I\!I}$ by

$$\delta_{\mathit{I\!I}} = \sup_{I \in \mathcal{I} \, : \, I \subseteq \mathit{I\!I}} \Delta_I / v_I .$$

The next assertion claims that the risk of the adaptive estimate is not larger in order than $v_{\mathit{I\!I}}$ for all $\mathit{I\!I}$ such that $\delta_{\mathit{I\!I}}$ is sufficiently small.

Theorem 2.3 *Let (2.8) hold true. Let an interval $\mathit{I\!I}$ be such that for some $D \ge 0$, it holds with a positive probability $\delta_{\mathit{I\!I}} \le D$. Then*

$$\begin{aligned} &\mathsf{P}\left(\mathit{I\!I} \text{ is rejected}, \, \delta_{\mathit{I\!I}} \le D \right) \\ &\le \sum_{I \in \mathcal{I}(\mathit{I\!I})} \sum_{J \in \mathcal{J}(I)} 12 \sqrt{e} \lambda_J (1 + \ln B) e^{-(\lambda_J - D)^2/(2a_\gamma)}, \end{aligned} \tag{2.11}$$

where $\lambda_J = \lambda(1 - s_\gamma \lambda N_J^{-1/2})$ with $N_J = \min\{|J|, |I \setminus J|\}$.

Moreover, if $N_J \ge 2 s_\gamma \lambda$ for all $J \in \mathcal{J}(I)$ and all $I \in \mathcal{I}$, then it holds for the adaptive estimate $\hat\theta = \tilde\theta_{\hat I}$ on the random set $A = \{\mathit{I\!I}$ is not rejected, $\delta_{\mathit{I\!I}} \le D\}$:

$$|\tilde\theta_I - \tilde\theta_{\mathit{I\!I}}| \le 2\lambda \tilde v_{\mathit{I\!I}}$$

and

$$|\tilde\theta_I - \theta_n| \le \left\{ D + 3\lambda + 2\lambda s_\gamma (D+\lambda) |\mathit{I\!I}|^{-1/2} \right\} v_{\mathit{I\!I}}.$$

Remark 2.2 *It is easy to see that the sum on the right hand side of the bound (2.11) can be made arbitrarily small by a proper choice of the constant* λ *and the sets* $\mathcal{J}(I)$. *Hence, the result of the theorem claims that with a dominating probability a "good" interval* $I\!I$ *will not be rejected and the adaptive estimate* $\hat{\theta}$ *is up to a constant factor as good as any of the "good" estimates* $\tilde{\theta}_{I\!I}$.

Remark 2.3 *As mentioned in Remark 2.1, the probability bound in the right hand side of (2.11) can refined for the special case when the process* θ_t *is constant within* $I\!I$ *by replacing the factor* $12\sqrt{e}\lambda_J(1+\ln B)e^{-(\lambda_J-D)^2/(2a_\gamma)}$ *with* $6e^{-\lambda_J^2/(2a_\gamma)}$.

2.5 Change point model

A *change point* model is described by a sequence $\nu_1 < \nu_2 < \ldots$ of stopping times with respect to the filtration $\mathcal{F}_t$ and by values $\sigma_{(1)}, \sigma_{(2)}, \ldots$ where each $\sigma_{(k)}$ is $\mathcal{F}_{\nu_k}$-measurable. By definition, $\sigma_t = \sigma_{(k)}$ for $\nu_k < t \le \nu_{k+1}$ and σ_t is constant for $t < \nu_1$. This is an important special case of the model (2.1). For this special case, the above procedure has a very natural interpretation: when estimating at the point n we search for a largest interval of the form $[n-m, n]$ that does not contain a change point. This is doing via testing for a change point within the interval candidate $I = [n-m, n]$. Note that the classical maximum likelihood test for no change point in the regression case with Gaussian $\mathrm{N}(0, \sigma^2)$ errors is also based on comparison of the mean values of observations Y_t over the subintervals $I = [n-m, n-m']$ and every subinterval $J = [n-m', n]$ for different m', so that the proposed procedure has strong appeal in this situation. However, there is an essential difference between testing of a change point and testing of homogeneity appearing as a building block of our adaptive procedure. Usually, a test for a change point is constructed in a way to provide the prescribed probability of a "false alarm", that is, rejecting the "no change point" hypothesis under homogeneity. Our adaptive procedure involves a lot of such tests for every candidate I, which leads to a multiple testing problem. As a consequence, each particular test

should be performed at a very high level; that is, it should be rather conservative providing a joint error probability at a reasonable level.

2.5.1 Probability of a "false alarm"

For the change point model, a "false alarm" would mean that the interval candidate I is rejected although the hypothesis of homogeneity is still fulfilled. The arguments used in the proof of Theorem 2.3 lead to the following upper bound for the probability of a "false alarm":

Theorem 2.4 *If $I = [n-m, n]$ is an interval of homogeneity, that is $\theta_t = \theta_n$ for all $t \in I$, then*

$$\mathsf{P}\,(I \text{ is rejected}) \le \sum_{I \in \mathcal{I}(I\!I)} \sum_{J \in \mathcal{J}(I)} 6 \exp\left\{ -\frac{\lambda^2}{2a_\gamma (1 + \lambda s_\gamma |J|^{-1/2})^2} \right\}.$$

This result is a special case of Theorem 2.3 with $\Delta_J \equiv 0$ when taking into account Remark 2.3.

Theorem 2.3 implies that for every fixed value M there exists a fixed λ providing a prescribed upper bound α for the probability of a "false alarm" for a homogeneous interval I of length M. Namely, the choice

$$\lambda \ge (1+\epsilon) \sqrt{2a_\gamma \ln \frac{M}{m_0 \alpha}}$$

leads for a proper small positive constant $\epsilon > 0$ to the inequality

$$\sum_{I \in \mathcal{I}(I\!I)} \sum_{J \in \mathcal{J}(I)} 6 \exp\left\{ -\frac{\lambda^2}{2a_\gamma (1 + \lambda s_\gamma |J|^{-1/2})^2} \right\} \le \alpha.$$

Here, M/m_0 is approximately the number of intervals in $\mathcal{J}(I)$ .This bound is, however, very rough, and it is only of theoretical importance since we estimate the probability of the sum of dependent events by the sum of single probabilities. The value of λ providing a prescribed probability of a "false alarm" can be found by Monte Carlo simulation for the homogeneous model with the constant volatility as described in Section 2.6.

2.5.2 Sensitivity to change points and the mean delay

The quality (sensitivity) of a change point procedure is usually measured by the mean delay between the occurrence of a change point and its detection. To study this property of the proposed method, we consider the case of estimation at a point n immediately after a change point ν. It is convenient to suppose that ν belongs to the end points of an interval which is tested for homogeneity. In this case the "ideal" choice $I\!I$ is clearly $[\nu, n]$. Theorem 2.3 claims that the quality of estimation at n is essentially the same as if we knew the latest change point ν a priori. In fact, one can state a slightly stronger assertion: every interval I which is essentially larger than $I\!I$ will be rejected with a high probability provided that the magnitude of the change is large enough.

Denote $m' = |I\!I|$, that is, $m' = n - \nu$. Let also $I = [\nu - m, n] = [n - m' - m, n]$ for some m, so that $|I| = m + m'$, and let θ (resp. θ') denote the value of the parameter θ_t before (resp. after) the change point ν. The magnitude of the change point is measured by the relative change $b = 2|\theta' - \theta|/\theta$.

It is worth mentioning that the values θ_t and especially θ'_t can be random and dependent on past observations. For instance, θ'_t may depend on Y_t for all $t < \nu$. The interval I will certainly be rejected if $|\tilde{\theta}_{I \setminus I\!I} - \tilde{\theta}_{I\!I}|$ is sufficiently large compared to the corresponding critical value.

Theorem 2.5 *Let* $\mathsf{E}(Y_t \mid \mathcal{F}_{t-1}) = \theta$ *before the change point at* ν *and* $\mathsf{E}(Y_t \mid \mathcal{F}_{t-1}) = \theta'$ *after it, and let* $b = |\theta' - \theta|/\theta$. *Let* $I = [n - m' - m, n]$ *with* $m' = n - \nu$. *If* $\varrho := \lambda s_\gamma / \sqrt{\min\{m, m'\}} < 1$ *and*

$$b \geq \frac{2\varrho + \sqrt{2}\varrho(1 + \varrho)}{1 - \varrho}, \tag{2.12}$$

then $\mathsf{P}(I \text{ is not rejected}) \leq 4e^{-\lambda^2/(2a_\gamma)}$.

The result of Theorem 2.5 delivers some additional information about the sensitivity of the proposed procedure to change points. One possible question is about the minimal delay m' between the change point ν and the first moment n when the procedure starts to indicate this change point by selecting an interval of type $I\!I = [\nu, n]$. Due to Theorem 2.5, the change will be "detected" with a high probability if the value $\varrho = \lambda s_\gamma / \sqrt{m'}$ fulfills (2.12). With fixed $b > 0$, condition (2.12) leads to $\varrho \le bC_0$ for some fixed constant C_0. The latter condition can be rewritten in the form $m' \ge b^{-2}\lambda^2 s_\gamma^2 / C_0^2$. We see that this lower bound for the required delay m' is proportional to b^{-2}, where b is the change point magnitude. It is also proportional to the threshold λ squared. In turn, for the prescribed probability α of rejecting a homogeneous interval of length M, the threshold λ can be bounded by $C\sqrt{\ln \frac{M}{m_0 \alpha}}$. In particular, if we fix the length M and α, then $m' = O(b^{-2})$. If we keep fixed the values b and M but aim to provide a very small probability of a "false alarm" by letting α go to zero, then $m' = O(\ln \alpha^{-1})$. All these issues are in agreement with the theory of change point detection; see, for example Pollak (1985) and Brodskij and Darkhovskij (1993).

2.6 LAVE in practice

The aim of this section is to give some hints concerning the choice of the testing intervals and the smoothing parameter λ and to illustrate the performance of the LAVE procedure on simulated and real data. We consider the simplest homogeneous model and we study the stability of the procedure in such a situation. Then a change point model is analyzed and the sensitivity with respect to the jump magnitude is measured. Finally, LAVE is applied to a set of exchange rate data.

2.6.1 Choice of the sets $\mathcal{I}$ and $\mathcal{J}(I)$

The presented algorithm involves the sets of interval candidates $\mathcal{I}$ and of testing intervals $\mathcal{J}(I)$. The simplest proposal is based on the use of a regular time grid $t_1, t_2, \ldots$, with grid step $m_0 \in \mathbb{N}$, that is, $t_k = m_0 k$, $k = 1, 2, \ldots$. For a given time point n, the set $\mathcal{I}$ of interval candidates is defined in the

following way:

$$\mathcal{I} = \{I_k = [t_k, n] \,:\, t_k \le n - m_0,\, k = 1, 2 \ldots\}$$

Next, for every interval I_k, we define the set $\mathcal{J}(I_k)$ of testing subintervals $J_{k'} \subset I_k$ such that $J_{k'} = [t_{k'}, n]$ for all $t_{k'} > t_k$ belonging to the grid. The homogeneity within I_k is then tested by comparing the pairs of estimates $\tilde{\theta}_J$ and $\tilde{\theta}_{I_k \setminus J}$ for all $J \in \mathcal{J}(I_k)$.

In this construction, the sets $\mathcal{I}$, $\mathcal{J}(I)$ are completely determined by the grid step m_0. The value of m_0 should be selected possibly small, because it represents the minimal delay before the LAVE algorithm can detect a change point. Nevertheless, m_0 should be sufficiently large to provide stability of the estimates $\tilde{v}_J$ and $\tilde{v}_{I \setminus J}$. For the simulation and the analysis of real data, we use $m_0 = 10$, which represents a good compromise. However, small changes in this value, that is, $5 \le m_0 \le 20$, do not appear to have great influence on the estimation results.

2.6.2 Choice of λ and γ

The selection of γ and in particular λ is more critical. Theorem 2.4 suggests that, in the context of a change point model, a reasonable approach for selecting λ is by providing a prescribed level α for rejecting a homogeneous interval I of a given length M. This would clearly imply at most the same level α for rejecting a homogeneous interval of a smaller length. However, the value of λ which can be derived with the help of Theorem 2.4 is rather conservative. A more accurate choice can be made by Monte Carlo simulation. We examine the procedure described in Section 2.3 with the sets of intervals $\mathcal{I}$ and $\mathcal{J}(I)$ on the regular grid with the fixed step $m_0 = 10$. A constant (and therefore also time homogeneous) model assumes that the parameter θ_t does not vary in time, that is, $\theta_t \equiv \theta$. It can easily be seen that the value θ has no influence on the procedure under time homogeneity. One can therefore suppose that $\theta = 1$ and the original model (2.1) is transformed into the regression model $Y_t = 1 + s_\gamma \zeta_t$ with the constant trend and homogeneous variance s_γ. This model is completely described and therefore,

one can determine by simulation the value of λ for which an interval of time homogeneity of length M is not rejected with a frequency of 95%.

The values of λ are computed for $M = 40$ and 80 and for the power transformations $\gamma = 0.5$, 1.0 and 2.0. The results are shown in Table 2.1. Note that the values of λ calibrated for $M = 80$ are necessarily larger and therefore more conservative than the values of λ calibrated for $M = 40$.

Table 2.1: The value of λ, which for a given power transformation γ provide the rejection of an interval of time homogeneity of length M with a frequency of 5%.

Smoothing Parameters					
$\gamma = 0.5$		$\gamma = 1.0$		$\gamma = 2.0$	
$M = 80$	$M = 40$	$M = 80$	$M = 40$	$M = 80$	$M = 40$
$\lambda = 2.74$	$\lambda = 2.40$	$\lambda = 2.58$	$\lambda = 2.24$	$\lambda = 2.18$	$\lambda = 1.86$

2.6.3 Simulation results for the change point model

We now evaluate the performance of the LAVE algorithm on simulated data. Two change point time series of length $T = 240$ are considered. The simulated data display two jumps of the same magnitude in opposite directions: $\sigma_t = \sigma$ for $t \in [1, 80]$ and $t \in [161, 240]$ and $\sigma_t = \sigma'$ for $t \in [81, 160]$, where $\sigma = 1$ and $\sigma' = 3$ and 5 respectively. For each model, 500 realizations are generated, and the estimation is performed at each time point $t \in [t_0, 240]$, where t_0 is set equal to 20.

We compute the estimation error for each combination of γ and λ with the following criterion:

$$\sum_{t=20}^{240} \sum_{\omega=1}^{500} \left(\frac{\hat{\sigma}_t - \sigma_t}{\sigma_t} \right)^2 (\omega), \tag{2.13}$$

where the index ω indicates the realizations of the change point model. We note that, in (2.13) the quadratic error is divided by the true volatility so

that the criterion does not depend on the scale of σ_t. The results shown in Table 2.2 are favorable to the choice of the smaller value of γ, confirming that the loss of efficiency caused by $\gamma < 2$ is offset by the greater normality of the errors. Figure 2.2 and 2.3 show the results of the estimation for the power

Table 2.2: Estimation errors for all the combinations of parameters γ and λ.

Estimation Error						
Parameters	$\gamma = 0.5$		$\gamma = 1.0$		$\gamma = 2.0$	
	$\lambda = 2.74$	$\lambda = 2.40$	$\lambda = 2.58$	$\lambda = 2.24$	$\lambda = 2.18$	$\lambda = 1.86$
Small Jump	19241.9	17175.3	19121.2	16522.5	24887.2	17490.9
Large Jump	46616.2	43282.5	51363.9	46706.4	68730.7	55706.3

transformation $\gamma = 0.5$ and the value of λ calibrated for an interval of time homogeneity of length $M = 40$ and $M = 80$ respectively. The plots on the top display the true process (straight line), the empirical median among all estimates (thick dotted line) and the empirical quartile among all estimates (thin dotted line). The plots of the bottom line similarly display the length of the interval of time homogeneity, which is minimal (resp. maximal) just after (resp. just before) a change point, and the median and the quartiles among all estimates.

The results are satisfactory. The volatility is estimated precisely and the change points are quickly detected. As expected, the behavior of the method within homogeneous regions is very stable. The delay in detecting a change point becomes smaller as the jump size grows. Taking a smaller λ also results in a smaller delay and improves the quality of estimation after the change points. The results for other power transformations look very similar and therefore are not displayed.

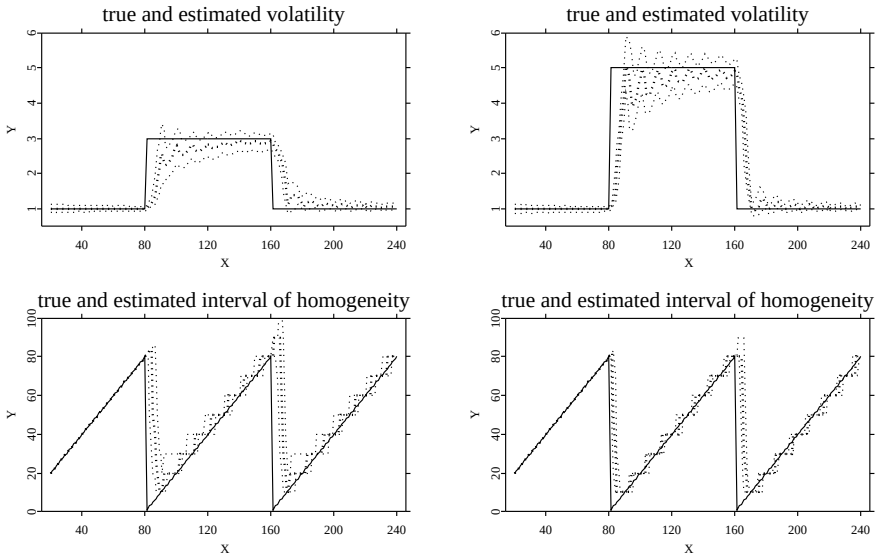

Figure 2.2: Estimation results for the change point model. The upper plots show the values of the standard deviation, while the lower plots show the values of the interval of homogeneity at each time point. True values (solid line), median of all estimates (thick dotted line), upper and lower quartiles (thin dotted line). The value of λ for $\gamma = 0.5$ and $M = 40$ has been used.

2.6.4 Estimation of exchange rate volatility

We apply the LAVE procedure to a set of nine exchange rates, which are available from the web page of the US Federal Reserve. The data sets represent daily exchange rates of the US Dollar (USD) against the following currencies: Australian Dollar (AUD), British Pound (BPD), Canadian Dollar (CAD), Danish Krone (DKR), Japanese Yen (JPY), Norwegian Krone (NKR), New Zealand Dollar (NZD), Swiss Franc (SFR) and Swedish Krone (SKR). The period under consideration goes from January 1, 1990, to April 7, 2000.

All the time series show qualitatively almost the same pattern, therefore we provide the graphical example only for the two representative exchange rates JPY/USD and BPD/USD (Figure 2.4). The empirical mean of the

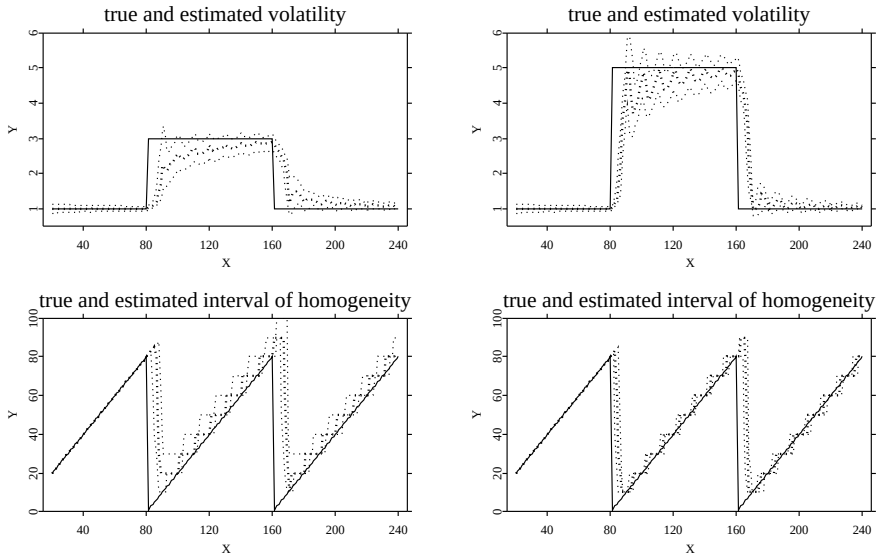

Figure 2.3: Estimation results for the change point model. The upper plots shows the values of the standard deviation, while the lower plots shows the values of the interval of homogeneity at each time point. True values (solid line), median of all estimates (thick dotted line), upper and lower quartile (thin dotted line). The value of λ for $\gamma = 0.5$ and $M = 80$ has been used.

Table 2.3: Summary statistics.

Currency	N	Mean$\cdot 10^5$	Variance$\cdot 10^5$	Skewness	Kurtosis
AUD	2583	-10.41	3.19	-0.18	8.85
BPD	2583	-0.67	3.53	-0.27	5.79
CAD	2583	8.81	0.89	0.04	5.49
DKR	2583	6.09	4.20	-0.03	4.96
JPY	2583	-12.70	5.48	-0.58	7.36
NKR	2583	9.49	4.25	0.31	8.63
NZD	2583	-6.58	3.60	-0.35	49.17
SFR	2583	1.48	5.40	-0.18	4.52
SKR	2583	12.66	4.61	0.37	9.66

returns is close to zero, while the empirical kurtosis is larger than 3. Furthermore, variance clustering and persistence of the autocorrelation of the square returns is also visible. The estimated standard deviation is nicely in accordance with the development of the volatility and, in particular, sharp changes in the volatility tend to be quickly recognized. Note also that the variability of the estimated interval of time homogeneity appears to grow as the estimated interval becomes larger. This is a feature of the algorithm because the number of tests grows with the accepted interval, so that a rejection becomes more probable. Nevertheless, this variability does not affect strongly the estimated volatility coefficient. Figure 2.5 shows the significantly persistent autocorrelation of the absolute returns, together with the autocorrelation of the absolute returns divided by the estimated standard deviation. The autocorrelation of the standardized absolute returns is not significant any more, and this fact supports the choice of a locally homogeneous model in order to explain the data.

A benchmark model

As a matter of comparison, we also consider a model which is commonly used to estimate and forecast volatility processes: the GARCH(1,1) model proposed by Bollerslev (1986):

$$\sigma_t^2 \;=\; \omega + \alpha R_{t-1}^2 + \beta \sigma_{t-1}^2.$$

Among all parametric volatility models, it represents the most common specification: *"The GARCH(1,1) is the leading generic model for almost all asset classes of returns. ... it is quite robust and does most of the work in almost all cases."* (Engle, 1995).

We do not require the parameters to be constant through the whole sample, but, similarly to Franses and Dijk (1996), we consider a rolling estimate. We thus fit the model to a sample of 350 observations, generate the forecast, delete the first observation from the sample and add the next one. Such a procedure reduces the harmful effect of possible parameter shifts on the forecasting performances of the model, even if at the same time it may increase

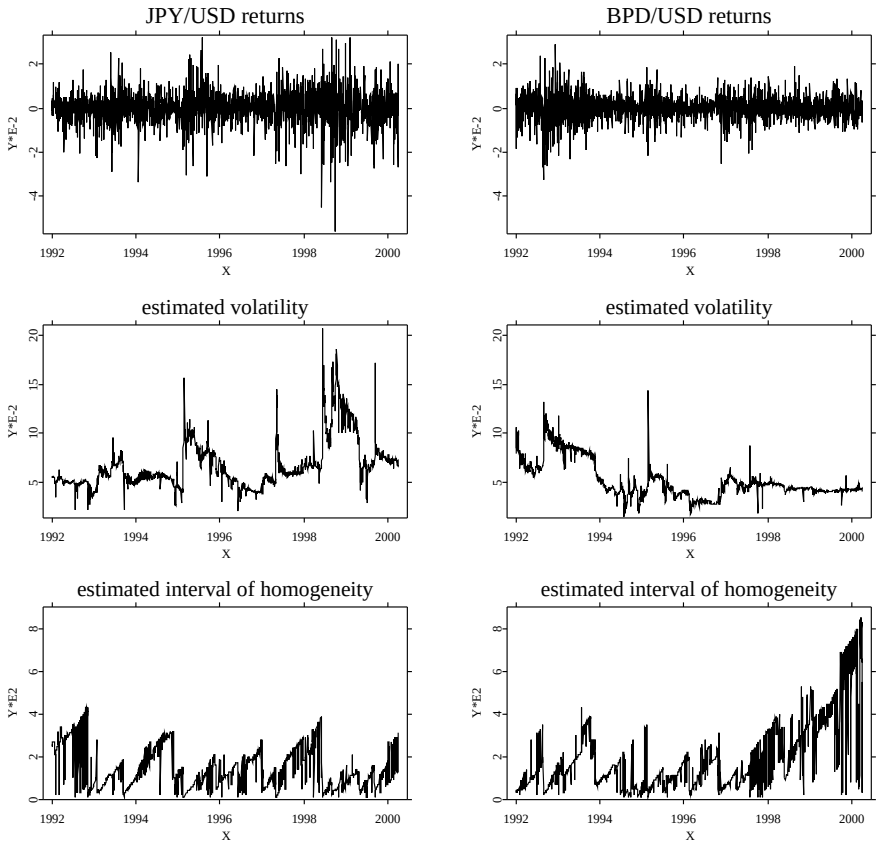

Figure 2.4: Exchange rate returns, estimated standard deviation and estimated interval of time homogeneity. The value of λ for $\gamma = 0.5$ and $M = 80$ has been used.

the estimation variability.

The volatility is a hidden process which can be observed only together with a multiplicative error; therefore, the evaluation of the forecasting performance of an algorithm is not straightforward. Due to the model (2.1), $\mathsf{E}\left(R_{t+1}^2 \mid \mathcal{F}_t\right) = \sigma_{t+1}^2$. Therefore, given a forecast $\hat{\sigma}_{t+1|t}$, the empirical mean value of $|R_{t+1}^2 - \hat{\sigma}_{t+1|t}^2|^p$ can be used to measure the quality of this forecast. The forecast ability of the LAVE and the GARCH estimates are therefore

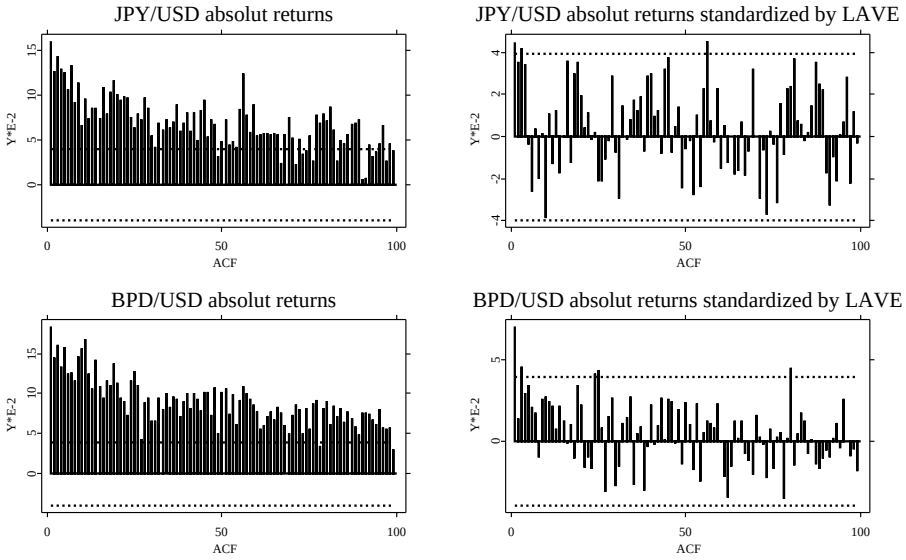

Figure 2.5: ACF of the absolute values of the exchange rate returns and ACF of the absolute values of the exchange rate returns standardized by LAVE

evaluated with the following criterion:

$$\frac{1}{T - t_0 - 1} \sum_{t=t_0}^{T} \left| R_{t+1}^2 - \hat{\sigma}_{t+1|t}^2 \right|^p \quad \text{with} \quad p = 0.5 \; .$$

The value of $p = 0.5$ is chosen instead of the more common $p = 2$ because we are interested in a robust criterion which is not too sensitive to the presence of outliers. The relative performance of the LAVE and the GARCH estimates is displayed in Table 2.4. The performance of the LAVE approach is clearly better; furthermore, the table gives a clear hint for the choice of the power transformation. Indeed, $\gamma = 0.5$ provides the smallest forecasting errors, while $\gamma = 2.0$ leads to the largest forecasting errors, which are sometimes larger than that of the GARCH model.

Table 2.4: Forecast performance of LAVE relative to GARCH.

Currency	$\gamma = 0.5$		$\gamma = 1.0$		$\gamma = 2.0$	
	$M = 80$	$M = 40$	$M = 80$	$M = 40$	$M = 80$	$M = 40$
AUD	0.94	0.94	0.96	0.96	0.99	0.98
BPD	0.96	0.96	0.97	0.97	1.00	1.00
CAD	0.97	0.97	0.98	0.99	1.01	0.99
DKR	0.97	0.98	0.98	0.98	1.01	1.00
JPY	0.95	0.94	0.97	0.96	1.00	0.99
NKR	0.96	0.95	0.97	0.96	0.99	0.98
NZD	0.87	0.87	0.90	0.90	0.95	0.94
SFR	0.98	0.98	0.99	0.99	1.00	1.00
SKR	0.96	0.96	0.97	0.96	0.98	0.97

2.7 Conclusions and outlook

The locally adaptive volatility estimate (LAVE) is described and analyzed in this paper. It provides a nonparametric way for estimating and short term forecasting the volatility of financial returns.

It is assumed that a local constant approximation of the volatility process holds over some unknown interval. The issue of filtering this interval of time homogeneity out of the return time series is considered, and a nonparametric approach is presented. The estimate of the volatility process is then found by averaging over the interval of time homogeneity.

A theoretical analysis of the properties of the LAVE algorithm is provided and the problem of selecting the smoothing parameters is analyzed through Monte Carlo simulation. The estimation results on change point models show that the method has a reasonable performance in practice. An empirical application to exchange rate returns and a comparison with a GARCH(1,1) also provides good evidence that the new method is competitive and can even outperform the standard parametric models especially for forecasting with a short horizon. An important feature of the proposed method is that

it allows for a straightforward extension on the multivariate volatility estimation; see Härdle et al. (2003) for a detailed discussion.

Obviously, if the underlying conditional distribution is not normal, the estimated volatility can give only partial information about the risk of the asset. Recent developments in risk analysis tends to focus on the estimation of the quantiles of the distribution. In this direction the LAVE procedure can be used as a convenient tool for pre-whitening the returns and obtain a sample of "almost" identical and independently distributed returns, which do not display any more variance clustering. So that the usual techniques of quantile estimation could be applied in a static framework. This application is discussed in detail in Chapter 4.

2.8 Appendix

In this section, we collect the proofs of the results stated above. We begin by considering some useful properties of the power transformation introduced in Section 2.2.

2.8.1 Some properties of the power transformation

Let $g_\gamma(u)$ be the moment generating function of $\zeta_\gamma = D_\gamma^{-1}\left(|\xi|^\gamma - C_\gamma\right)$:

$$g_\gamma(u) = \mathsf{E}e^{u\zeta_\gamma}\,.$$

It is easy to see that this function is finite for $\gamma < 2$ and all u and for $\gamma = 2$ and $u < 1$. For $\gamma = 1/2$, the function $2u^{-2}\ln g_\gamma(u)$ is plotted in Figure 2.6.

Lemma 2.1 *For every $\gamma \leq 1$, there exists a constant $a_\gamma > 0$ such that*

$$\ln \mathsf{E}e^{u\zeta_\gamma} \leq \frac{a_\gamma u^2}{2}. \tag{2.14}$$

(2.15)

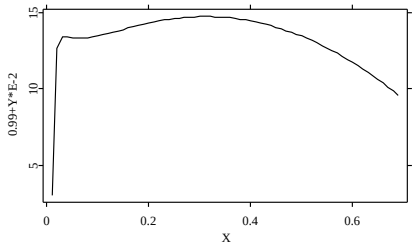

Figure 2.6: The log-Laplace transform of $\zeta_{1/2}$ divided by the log-Laplace transform of a standard normal r.v.

Proof It is easy to check that the function $g_\gamma(u)$ with $\gamma \le 1$ is positive and smooth (infinitely many times differentiable). Moreover, the function $h_\gamma(u) = \ln g_\gamma(u)$ is also smooth and satisfies $h_\gamma(0) = h'_\gamma(0) = 0$, $h''_\gamma(0) = \mathsf{E}\zeta_\gamma^2 = 1$. This yields that $u^{-2}h_\gamma(u) = u^{-2}\ln g_\gamma(u)$ is bounded on every finite interval of the positive semi-axis $[0,\infty)$. It therefore remains to show that

$$\lim_{u\to\infty} u^{-2}\ln \mathsf{E}e^{u\zeta_\gamma} < \infty.$$

Since $\zeta_\gamma(u) = D_\gamma^{-1}(|\xi|^\gamma - C_\gamma)$, it suffices to bound $u^{-2}\mathsf{E}e^{u|\xi|^\gamma/D_\gamma}$. For every $t > 0$

$$\begin{aligned}
\mathsf{E}e^{u|\xi|^\gamma D_\gamma^{-1}} &= \mathsf{E}e^{u|\xi|^\gamma D_\gamma^{-1}}\mathbf{1}(|\xi|\le t) + \mathsf{E}e^{u|\xi|^\gamma D_\gamma^{-1}}\mathbf{1}(|\xi|>t)\\
&\le e^{ut^\gamma D_\gamma^{-1}} + \mathsf{E}e^{u|\xi|t^{\gamma-1}D_\gamma^{-1}}\\
&\le e^{ut^\gamma D_\gamma^{-1}} + 2\mathsf{E}e^{u\xi t^{\gamma-1}D_\gamma^{-1}}\\
&= e^{ut^\gamma D_\gamma^{-1}} + 2e^{u^2t^{2\gamma-2}D_\gamma^{-2}}.
\end{aligned}$$

Next, with $t = u^{1/(2\gamma)}$ and $\gamma < 1$, it holds for $u\to\infty$:

$$\begin{aligned}
u^{-2}\ln e^{ut^\gamma D_\gamma^{-1}} &= u^{-1/2}D_\gamma^{-1} \to 0,\\
u^{-2}\ln e^{u^2t^{2\gamma-2}D_\gamma^{-2}} &= u^{-(1-\gamma)/\gamma}D_\gamma^{-2} \to 0.
\end{aligned}$$

For $\gamma = 1$, the last expression remains bounded and the assertion follows.

For $\gamma = 1/2$, condition (2.14) meets with $a_\gamma = 1.005$. The next technical statement is a direct consequence of Lemma 2.1.

Lemma 2.2 *Let c_t be a predictable process w.r.t. the filtration $\mathcal{F} = (\mathcal{F}_t)$, that is, every c_t is a function of previous observations $R_1, \dots, R_{t-1}$: $c_t = c_t(R_1, \dots, R_{t-1})$. Then the process*

$$\mathcal{E}_t = \exp\left(\sum_{s=1}^{t} c_s\zeta_s - \frac{a_\gamma}{2}\sum_{s=1}^{t} c_s^2\right)$$

is a supermartingale, that is,

$$\mathsf{E}\left(\mathcal{E}_t \mid \mathcal{F}_{t-1}\right) \le \mathcal{E}_{t-1}. \tag{2.16}$$

(2.17)

The next result has been stated in Lipster and Spokoiny (1999) for Gaussian martingales, however, the proof is based only on the property (2.16) and allows for a straightforward extension to the sums of the form $M_t = \sum_{s=1}^{t} c_s\zeta_s$.

Theorem 2.6 *Let $M_t = \sum_{s=1}^{t} c_s\zeta_s$ with predictable coefficients c_s. Let then T be fixed or a stopping time. For every $b > 0$, $B \ge 1$ and $\lambda \ge 1$*

$$\mathsf{P}\left(|M_T| > \lambda\sqrt{\langle M\rangle_T},\ b \le \sqrt{\langle M\rangle_T} \le bB\right) \le 4\sqrt{e}\lambda\,(1+\ln B)\,e^{-\lambda^2/(2a_\gamma)}$$

where

$$\langle M\rangle_T = \sum_{t=1}^{T} c_t^2. \tag{2.18}$$

Remark 2.4 *If the coefficients c_t are deterministic or independent of M then Lemma 2.1 and the Chebyshev inequality yield*

$$\mathsf{P}\left(|M_T| > \lambda\sqrt{\langle M\rangle_T}\,\right) \le 2e^{-\lambda^2/(2a_\gamma)}. \tag{2.19}$$

2.8.2 Proof of Theorem 2.1

Define

$$\bar{\theta}_I = \frac{1}{|I|}\sum_{t\in I}\theta_t \qquad \xi_I = s_\gamma |I|^{-1}\sum_{t\in I}\theta_t\zeta_t\,.$$

Then $\tilde{\theta}_I = \bar{\theta}_I + \xi_I$. By the definition of Δ_I

$$|\bar{\theta}_I - \theta_n| = |I|^{-1}\left|\sum_{t\in I}(\theta_t - \theta_n)\right| \le \Delta_I \tag{2.20}$$

Next, by (2.5)

$$\tilde{\theta}_I - \theta_n = \bar{\theta}_I - \theta_n + \xi_I$$

and the use of (2.20) yields

$$\mathsf{P}\left(|\tilde{\theta}_I - \theta_n| > \Delta_I + \lambda v_I\right) \le \mathsf{P}\left\{\left|\sum_{t\in I}\theta_t\zeta_t\right| > \lambda\left(\sum_{t\in I}\theta_t^2\right)^{1/2}\right\}.$$

In addition, if the volatility coefficient σ_t satisfies $b \le \sigma_t^2 \le bB$ with some positive constant b, B, then the conditional variance $v_I^2 = s_\gamma^2|I|^{-2}\sum_{t\in I}\theta_t^2$ fulfills

$$b'|I|^{-1} \le v_I^2 \le b'|I|^{-1}B$$

with $b' = bs_\gamma^2$. Now the assertion follows from (2.8) and Theorem 2.6.

2.8.3 Proof of Theorem 2.2

It suffices to show that the inequalities $\Delta_I/v_I \le D$ and

$$|\tilde{\xi}_I| = |\tilde{\theta}_I - \bar{\theta}_I| \le \lambda v_I \tag{2.21}$$

imply $|\tilde{\theta}_I - \theta_n| \le \lambda'\tilde{v}_I$ where λ' solves the equation $D + \lambda = \lambda'/(1 + \lambda' s_\gamma |I|^{-1/2})$. This would yield the desirable result by Theorem 2.6, cf. the proof of Theorem 2.1.

Lemma 2.3 *Let* $(\Delta_I/v_I)s_\gamma|I|^{-1/2} < 1$. *Under (2.21)*

$$\tilde{v}_I \geq v_I \left\{ \sqrt{1-(\Delta_I/v_I)^2 s_\gamma^2 |I|^{-1}} - s_\gamma \lambda |I|^{-1/2} \right\} \geq v_I \left\{ 1 - s_\gamma |I|^{-1/2} (\Delta_I/v_I + \lambda) \right\}.$$

Proof By definition of $\tilde{v}_I$ in view of (2.21)

$$\tilde{v}_I = s_\gamma \tilde{\theta}_I |I|^{-1/2} \geq s_\gamma \left(\bar{\theta}_I - \lambda v_I \right) |I|^{-1/2}.$$

Since $\bar{\theta}_I$ is the arithmetic mean of θ_t over I ,

$$\sum_{t \in I} (\theta_t - \bar{\theta}_I)^2 \leq \sum_{t \in I} (\theta_t - \theta_n)^2 \leq \Delta_I^2 |I|.$$

Next,

$$s_\gamma^{-2} |I| v_I^2 = |I|^{-1} \sum_{t \in I} \theta_t^2 = \bar{\theta}_I^2 + |I|^{-1} \sum_{t \in I} (\theta_t - \bar{\theta}_I)^2 \leq \bar{\theta}_I^2 + \Delta_I^2 ,$$

so that

$$\bar{\theta}_I \geq s_\gamma^{-1} |I|^{1/2} v_I \sqrt{1 - (\Delta_I s_\gamma v_I^{-1} |I|^{-1/2})^2} .$$

Hence, under (2.21)

$$\tilde{v}_I \geq v_I \left\{ \sqrt{1 - (\Delta_I s_\gamma v_I^{-1} |I|^{-1/2})^2} - s_\gamma \lambda |I|^{-1/2} \right\}$$

and the assertion follows.

The bound (2.21) and the definition of Δ_I imply

$$|\tilde{\theta}_I - \theta_n| \leq |\bar{\theta}_I - \theta_n| + |\tilde{\theta}_I - \bar{\theta}_I| \leq \Delta_I + \lambda v_I \leq (D + \lambda) v_I .$$

By Lemma 2.3 $\tilde{v}_I \geq v_I \left(1 - s_\gamma D |I|^{-1/2} - s_\gamma \lambda |I|^{-1/2} \right)$. Thus,

$$|\tilde{\theta}_I - \theta_n| \leq \frac{D + \lambda}{1 - s_\gamma (D + \lambda) |I|^{-1/2}} \, \tilde{v}_I = \lambda' \tilde{v}_I$$

as required.

2.8.4 Proof of Theorem 2.3

Let $I\!I$ be a "good" interval in the sense that with a high probability, $\Delta_J/v_J \le D$ for some nonnegative constant D and every $J \in \mathcal{J}(I\!I)$. First we show that $I\!I$ will not be rejected with a high probability provided that λ is sufficiently large.

We proceed similarly to the proof of Theorems 2.1 and 2.2. The procedure involves the estimates $\tilde\theta_J$, $\tilde\theta_{I\setminus J}$ and the differences $\tilde\theta_J - \tilde\theta_{I\setminus J}$ for all $I \in \mathcal{I}(I\!I)$ and all $J \in \mathcal{J}(I)$. The expansion $\tilde\theta_J = \bar\theta_J + \xi_J$ implies

$$\tilde\theta_J - \tilde\theta_{I\setminus J} = \bigl(\bar\theta_J - \bar\theta_{I\setminus J}\bigr) + \bigl(\xi_J - \xi_{I\setminus J}\bigr)\,.$$

Under the condition $\delta_I \le D$

$$|\bar\theta_J - \bar\theta_{I\setminus J}| \le \Delta_I \le D v_I \le D\sqrt{v_J^2 + v_{I\setminus J}^2}.$$

Define the events

$$\begin{aligned} A_I &= \bigcup_{J\in\mathcal{J}(I)} \left\{ |\xi_J - \xi_{I\setminus J}| \le (\lambda_J - D)\sqrt{v_J^2 + v_{I\setminus J}^2} \text{ and } \sqrt{\frac{\tilde v_J^2 + \tilde v_{I\setminus J}^2}{v_J^2 + v_{I\setminus J}^2}} \ge 1 - s_\gamma \lambda N_J^{-1/2} \right\} \\ A_{I\!I} &= \bigcup_{I\in\mathcal{I}: I\subseteq I\!I} A_I \end{aligned}$$

where $N_J = \min\{|J|, |I\setminus J|\}$ and $\lambda_J = \lambda(1 - s_\gamma\lambda N_J^{-1/2})$.

Define $A_{I\!I}^* = A_{I\!I} \cap \{\delta_{I\!I} \le D\}$. On this set

$$\begin{aligned} \frac{|\tilde\theta_J - \tilde\theta_{I\setminus J}|}{\sqrt{\tilde v_J^2 + \tilde v_{I\setminus J}^2}} &\le \frac{|\bar\theta_J - \bar\theta_{I\setminus J}| + |\xi_J - \xi_{I\setminus J}|}{\sqrt{\tilde v_J^2 + \tilde v_{I\setminus J}^2}} \\ &\le (D + \lambda_J - D)\sqrt{\frac{v_J^2 + v_{I\setminus J}^2}{\tilde v_J^2 + \tilde v_{I\setminus J}^2}} \le \frac{D + \lambda_J - D}{1 - s_\gamma\lambda N_J^{-1/2}} = \lambda. \end{aligned}$$

It is easy to see that the conditional variance of $\xi_J - \xi_{I\setminus J}$ is equal to $v_J^2 + v_{I\setminus J}^2$. Arguing similarly to Lemma 2.3 and Theorem 2.1 we bound with $\lambda_{J,D} =$

$\lambda_J - D$

$$
\begin{aligned}
&\mathsf{P}(A_I) \\
&\le \sum_{J \in \mathcal{J}(I)} \mathsf{P}\left(\frac{|\xi_J|}{v_J} > \lambda_{J,D} \right) + \mathsf{P}\left(\frac{|\xi_{I\setminus J}|}{v_{I\setminus J}} > \lambda_{J,D} \right) + \mathsf{P}\left(\frac{|\xi_J - \xi_{I\setminus J}|}{\sqrt{v_J^2 + v_{I\setminus J}^2}} > \lambda_{J,D} \right) \\
&\le \sum_{J \in \mathcal{J}(I)} 12\sqrt{e}\lambda_J (1 + \ln B) e^{-\lambda_{J,D}^2/(2a_\gamma)}
\end{aligned}
$$

and the first assertion of the theorem follows.

Now we show that on the set $A_{I\!I}^*$ the estimate $\hat\theta = \tilde\theta_{\hat I}$ fulfills $|\hat\theta - \hat\theta_{I\!I}| \le 2\lambda\tilde v_{I\!I}$. Due to the above, on $A_{I\!I}^*$ the interval $I\!I$ will not be rejected and hence, $|\hat I| \ge |I\!I|$. Let I be an arbitrary interval from $\mathcal{I}$ which is not rejected by the procedure. By construction, the $I\!I$ is one of the testing intervals for I. Denote $J = I \setminus I\!I$. Note that $|I|(\tilde\theta_I - \tilde\theta_{I\!I}) = |J|(\tilde\theta_J - \tilde\theta_{I\!I})$, so that the event "$I$ is not rejected" implies $|\tilde\theta_J - \tilde\theta_{I\!I}| \le \lambda\sqrt{\tilde v_J^2 + \tilde v_{I\!I}^2}$ and

$$
|\tilde\theta_I - \tilde\theta_{I\!I}| \le \frac{\lambda|J|}{|I|}\sqrt{\tilde v_J^2 + \tilde v_{I\!I}^2} \le \frac{\lambda|J|}{|I|}\,(\tilde v_J + \tilde v_{I\!I})\,.
$$

The use of $\tilde v_J = s_\gamma \tilde\theta_J |J|^{-1/2}$ and $|\tilde\theta_I - \tilde\theta_{I\!I}| \le \lambda\,(\tilde v_J + \tilde v_{I\!I})$ yields

$$
\bigl|\tilde v_J |J|^{1/2} - \tilde v_{I\!I} |I\!I|^{1/2}\bigr| \le \lambda s_\gamma\,(\tilde v_J + \tilde v_{I\!I})
$$

implying

$$
\tilde v_J \le \frac{|I\!I|^{1/2} + \lambda s_\gamma}{|J|^{1/2} - \lambda s_\gamma}\,\tilde v_{I\!I}\,, \qquad \tilde v_J + \tilde v_{I\!I} \le \frac{|J|^{1/2} + |I\!I|^{1/2}}{|J|^{1/2} - \lambda s_\gamma}\,\tilde v_{I\!I}\,.
$$

Therefore,

$$
|\tilde\theta_I - \tilde\theta_{I\!I}| \le \frac{\lambda|J|\,\bigl(|J|^{1/2} + |I\!I|^{1/2}\bigr)}{(|J| + |I\!I|)\,(|J|^{1/2} - \lambda s_\gamma)}\,\tilde v_{I\!I}
$$

It is straightforward to check that the function $f(x) = x^2(x+1)/\{(x^2+1)(x-c)\}$ with any $c \ge 0$ satisfies $f(x) \le 2$ for all $x \ge 2c$. This implies with $x = |J|^{1/2}/|I\!I|^{1/2}$ and $c = \lambda s_\gamma/|I\!I|^{1/2}$ that

$$
|\tilde\theta_I - \tilde\theta_{I\!I}| \le 2\lambda\tilde v_{I\!I}
$$

under the condition that $|J|^{1/2} \geq 2\lambda s_\gamma$.

Let $\Delta_{I\!I} \leq D v_{I\!I}$. Similarly to Lemma 2.3 $\tilde{v}_{I\!I} \leq v_{I\!I} \left\{1 + s_\gamma (D+\lambda) |I\!I|^{-1/2}\right\}$ and by Theorem 2.1 $|\tilde{\theta}_{I\!I} - \theta_n| \leq (D+\lambda) v_{I\!I}$. This yields

$$|\tilde{\theta}_I - \tilde{\theta}_{I\!I}| \leq 2\lambda v_{I\!I} \left\{1 + s_\gamma (D+\lambda) |I\!I|^{-1/2}\right\}$$

and

$$\begin{aligned} |\tilde{\theta}_I - \theta_n| &\leq 2\lambda v_{I\!I} \left\{1 + s_\gamma (D+\lambda) |I\!I|^{-1/2}\right\} + (D+\lambda) v_{I\!I} \\ &= \left\{D + 3\lambda + 2\lambda s_\gamma (D+\lambda) |I\!I|^{-1/2}\right\} v_{I\!I} \end{aligned}$$

as required.

2.8.5 Proof of Theorem 2.5

To simplify the exposition, we suppose that $\theta = 1$. (This does not restrict generality since one can always normalize each "observation" Y_t by θ .) We also suppose that $\theta' > 1$ and $b = \theta' - 1$. (The case when $\theta' < \theta$ can be considered similarly.) Finally we assume that $m' = m$ (One can easily see that this case is the most difficult one.) We again apply the decomposition

$$\tilde{\theta}_J = 1 + \xi_J \,, \qquad \tilde{\theta}_{I\!I} = \theta' + \xi_{I\!I}$$

see the proof of Theorem 2.1. Hence,

$$\tilde{\theta}_{I\!I} - \tilde{\theta}_J = b + \xi_{I\!I} - \xi_J \,.$$

It is straightforward to see that $v_J^2 = s_\gamma^2 / m$ and $v_{I\!I}^2 = s_\gamma^2 \theta' / m$. By Lemma 2.1 (see also Remark 2.4)

$$\mathsf{P}\left(|\xi_J| > \lambda v_J\right) + \mathsf{P}\left(|\xi'| > \lambda v_{I\!I}\right) \leq 4 e^{-\frac{\lambda^2}{2a\gamma}}$$

and it suffices to check that the inequalities $|\xi_J| \leq \lambda v_J$, $|\xi_{I\!I}| \leq \lambda v_{I\!I}$ and (2.12) imply

$$|\tilde{\theta}_J - \tilde{\theta}_{I\!I}| \geq \lambda \sqrt{\tilde{v}_J^2 + \tilde{v}_{I\!I}^2} \,.$$

Since $\theta' - 1 = b$ and since $\tilde{v}_J = s_\gamma |J|^{-1/2} \tilde{\theta}_J$ and similarly for $\tilde{v}_I$, one has, under the conditions $|\xi_J| \le \lambda v_J$, $|\xi_{I\!I}| \le \lambda v_{I\!I}$:

$$\begin{aligned} |\tilde{\theta}_J - \tilde{\theta}_{I\!I}| &\ge b - \frac{\lambda s_\gamma(\theta'+1)}{\sqrt{m}} = b(1-\varrho) - 2\varrho, \\ \tilde{v}_J &= \frac{s_\gamma}{\sqrt{m}}(1+\xi_J) \le \lambda^{-1}\varrho\,(1+\varrho)\,, \\ \tilde{v}_{I\!I} &= \frac{s_\gamma}{\sqrt{m}}(1+\xi_{I\!I}) \le \lambda^{-1}\varrho\,(1+\varrho) \end{aligned}$$

with $\varrho = m^{-1/2}\lambda s_\gamma$. Therefore

$$|\tilde{\theta}_J - \tilde{\theta}_{I\!I}| - \lambda\sqrt{\tilde{v}_J^2 + \tilde{v}_{I\!I}^2} \ge b(1-\delta) - 2\varrho - \sqrt{2}\varrho(1+\varrho) > 0$$

in view of (2.12) and the assertion follows.

Chapter 3

Estimation and Arbitrage Opportunities for Exchange Rate Baskets

3.1 Introduction

An exchange rate basket is a form of pegged exchange rate regime and it takes place whenever the domestic currency can be expressed as a linear combination of foreign currencies. Currency baskets are adopted by developing and transition countries in order to obtain a nominal anchor for monetary policy and are consistent with maintaining or demanding flexibility with respect to fluctuations among the exchange rates of the major international currencies. Recent crises involving emerging market economies have led some authors, e.g. Eichengreen et al. (1999), to conclude that pegged exchange rate regimes have an intrinsic tendency of developing crisis situations. Others, e.g. Mussa et al. (2000), do not agree with this view because these regimes have been successfully implemented by many countries. Nevertheless they confirm that pegged exchange rates may become an important source of vulnerability. In particular, modern capital mobility enables investors to exploit interest rate differentials which may arise between the domestic and the foreign currencies. Furthermore, if it is known that the monetary authorities are committed to sustain the exchange rate, such speculation seems virtually riskless and can

threaten the stability of the exchange rate regime.

The aim of this chapter is to analyze short -term portfolio investments, both from a finance-theoretical and an empirical viewpoint, in a capital market where an exchange rate is defined as a currency basket. In Section 3.2 the problem of investing in a currency basket is analyzed. A self financed investment strategy is developed which minimizes the expected quadratic cost function. The approach is inspired by mean variance hedging (Schäl, 1994) and provides a simple explicit solution for a problem of imperfect hedging. Along with the optimal hedging strategy an expression is also derived for the expected profit and for the expected variance of the profit. The estimate of the basket weights is required for the construction of the investment strategy, while the expected variance of the profit is needed in order to make an assessment about the risk of the investment.

The problems related with the estimation of these quantities are discussed in Section 3.3. From a statistical point of view, the currency basket regime can be represented by a regression model with stochastic regressors and possibly time varying coefficients. This estimation problem is also central to Christoffersen and Giorgianni (2000). The authors document on the superiority of the time varying parameters estimates, which are performed by means of Kalman filtering techniques. Kalman filters essentially assume a diffusion model for the time varying coefficient, which is suitable to describe continuous smooth movements in the weights. However, given that the Central Bank is more likely to intervene on the basket from time to time, the weights are more likely to jump occasionally and the estimation procedure has to account for this feature. To this end, an adaptive estimation algorithm is proposed based on a result by Lipster and Spokoiny (1999). The adaptive estimator rests on the assumption that the coefficients can be well approximated by a constant over some time interval. This feature is called *local time homogeneity*. The estimation strategy consists in detecting this *interval of time homogeneity* and then estimating the parameters over this interval with standard techniques, such as ordinary least squares. The adaptive estimator is a nonparametric technique because it does not require that the underlying

process belong to any specific parametric family, such as autoregressive or moving average processes. The only requirement is the local time homogeneity. This hypothesis is fulfilled particularly well by jump processes, which are constant over a certain interval and make up or down jumps at random times. A simulation study illustrates the performances of the new methodology for jump processes with jumps of different magnitude.

Finally, in Section 3.4 the optimal investment strategy is applied to the case of the Thai Baht basket. The basket weights are computed with the adaptive estimator. Furthermore we also implement a recursive estimator, a rolling estimator and the Kalman filter which serve as benchmark models. The expected and realized profits are calculated and the performances of the different estimators compared with profit based criteria.

The last section concludes and the appendix shows the equivalence between the mean variance hedging strategy and the risk minimizing strategy proposed by Christoffersen and Giorgianni (2000).

3.2 Investing in an exchange rate basket

An exchange rate basket regime takes place whenever a currency Y_t can be written as a linear combination of K other currencies. Taking the currency 1 as numeraire, i.e. $X_{1,t} \equiv 1$, one can express the value of the basket by the following equation:

$$Y_t = \sum_{j=1}^{K} \alpha_{j,t} X_{j,t} + \xi_t, \tag{3.1}$$

where $X_{j,t}$ is the amount of currency 1 per unit of currency j, i.e. the cross currency exchange rate. The stationary error term ξ_t, with $\mathsf{E}\xi_t = 0$ and $\mathsf{E}\xi_t^2 = \sigma^2$, models the fact that the above relationship usually holds only on average, because the exchange rate cannot be controlled exactly. Moreover the weights $\alpha_{j,t}$ may also change over time because of the reaction of the monetary authorities to changes in macroeconomic fundamentals such as trading patterns, and/or speculative pressures. Note that the central bank does not necessarily disclose the values of the basket weights to the public in

order to gain greater discretion in setting monetary policy.

The aim of this study is to analyze the possible strategies of an investor who wants to speculate on interest rate differentials which may arise among the countries whose currencies are in the basket. Suppose for simplicity that the country with currency Y_t has the highest interest rates. Then the main idea is to lend currency Y_t and to borrow a portfolio of the other currencies $X_{j,t}$, for $j = 1, \ldots, K$. Such an investment can have a positive expected profit, but it carries an intrinsic risk, so that we have to construct our investment strategy in order to minimize it. The *mean-variance hedging* strategy may provide a solution. This approach has been developed for the optimal hedging of non-attainable contingent claims and it focuses on the minimization of the tracking error at the terminal date, see e.g. Schäl (1994).

3.2.1 Mean-variance hedging

Before deriving the optimal investment strategy, some technical conditions are listed. In the remainder, it is assumed that the random variables: Y_t, $X_{j,t}$, $\alpha_{j,t}$ and ξ_t have finite second moments. The agents only observe the exchange rates Y_t and $X_{j,t}$ and the interest rates r_0 (for home currency deposits) and r_j, $j = 1, \ldots, K$ (for deposits in currencies composing the basket). Furthermore, they know that the central bank has committed itself to control the magnitude of the fluctuations of the home currency around a basket of known foreign currencies, whereby the values of the basket weights are not disclosed to the public. Therefore the information set of the agents only includes the past values of the exchange rates and interest rates. Finally it is assumed that ξ_{t+h}, $\alpha_{j,t+h}$ and $X_{j,t+h}$ for $j = 1, \ldots, K$ are conditionally independent given the observations up to time t. No distributional assumption is required on the ξ_t for the development of the financial strategy, while for the sake of the estimation we assume normality.

A two stage model is considered, where an investment decision is made at time t and the position is kept until time $t + h$. In the section devoted to the empirical analysis holding periods of 30 and 90 days, respectively will

be considered. This represents a simplification. Nevertheless, it is required because the model is estimated using inter-bank interest rates. Therefore, our assets consist of bank deposits and cannot be traded until maturity.

The mean-variance hedging problem can be formulated as follows: we have to determine at time t a strategy $(\psi_1, \ldots, \psi_K)$ such that the expected squared deviation of $\sum \psi_j X_{j,t+h}$ from one unit value of currency Y_{t+h} is minimized. Therefore the following quadratic cost function is considered:

$$\mathsf{E}_t \left\{ \left(Y_{t+h} - \textstyle\sum_{j=1}^{K} \psi_j X_{j,t+h} \right)^2 \right\}. \tag{3.2}$$

Substituting (3.1) for Y_{t+h} one gets:

$$\begin{aligned} & \mathsf{E}_t \left\{ \left(\textstyle\sum_{j=1}^{K} (\alpha_{j,t+h} - \psi_j) X_{j,t+h} + \xi_{t+h} \right)^2 \right\} \\ = \quad & \mathsf{E}_t \left\{ \left(\textstyle\sum_{j=1}^{K} (\mathsf{E}_t(\alpha_{j,t+h}) - \psi_j) X_{j,t+h} + \zeta_{t+h} \right)^2 \right\} \end{aligned}$$

where: $\zeta_{t+h} = \xi_{t+h} + \sum_{j=1}^{K} (\alpha_{j,t+h} - \mathsf{E}_t(\alpha_{j,t+h})) X_{j,t+h}$. The cross term is zero by the assumption of conditional independence and $\mathsf{E}_t(\zeta_{t+h}^2)$ does not depend on ψ_j for $j = 1, \ldots, K$. Differentiating with respect to ψ_j, for $j = 1, \ldots, K$, we obtain the following system of K first order conditions:

$$-\mathsf{E}_t \left[X_{j,t+h} \left\{ \textstyle\sum_{j=1}^{K} (\mathsf{E}_t(\alpha_{j,t+h}) - \psi_j^*) X_{j,t+h} \right\} \right] = 0, \qquad \text{for } j = 1, \ldots, K;$$

whose solution provides the optimal strategy: $\psi_j^* = \mathsf{E}_t(\alpha_{j,t+h})$ for $j = 1, \ldots, K$. This optimal strategy implies that the expected value of revising the hedging portfolio is zero:

$$\mathsf{E}_t \left(Y_{t+h} - \textstyle\sum_{j=1}^{K} \psi_j^* X_{j,t+h} \right) = \mathsf{E}_t(\zeta_{t+h}^2) = 0$$

and the expected quadratic costs are:

$$\begin{aligned} & \mathsf{E}_t \left\{ \left(Y_{t+h} - \textstyle\sum_{j=1}^{K} \psi_j^* X_{j,t+h} \right)^2 \right\} \\ = \quad & \sigma^2 + \mathsf{E}_t \left\{ \left(\textstyle\sum_{j=1}^{K} (\alpha_{j,t+h} - \mathsf{E}_t(\alpha_{j,t+h})) X_{j,t+h} \right)^2 \right\} \end{aligned} \tag{3.3}$$

Therefore, $(1+r_0)^{-1}Y_t$ is the amount of money which is needed at time t in order to hedge the portfolio $\sum \mathsf{E}_t(\alpha_{j,t+h})X_{j,t+h}$ in the mean-variance sense, where r_0 is the interest rate paid on a h day deposit in currency Y_t. On the other hand, if one prefers avoiding any risk one can simply buy at time t the discounted value of the portfolio: $\sum(1+r_j)^{-1}\mathsf{E}_t(\alpha_{j,t+h})X_{j,t}$. Therefore, the implementation of the mean-variance hedging strategy has positive expected profits if the following inequality holds:

$$(1+r_0)^{-1}Y_t < \textstyle\sum_{j=1}^{K}(1+r_j)^{-1}\mathsf{E}(\alpha_{j,t+h}|\mathcal{F}_t)X_{j,t}. \tag{3.4}$$

This means that an investment in the home currency is more convenient than an investment with the same expected revenue in the currencies composing the basket. In the empirical analysis of the Thai Baht basket one finds that the relationship (3.4) is fulfilled during the whole period under investigation for any of the four methods that are used to estimate the basket weights. If inequality (3.4) holds, then the speculative strategy can be implemented in the following way: first, one can borrow the portfolio $\sum(1+r_j)^{-1}\mathsf{E}_t(\alpha_{j,t+h})X_{j,t}$, lend the amount $(1+r_0)^{-1}Y_t$ and invest the difference $\sum(1+r_j)^{-1}\mathsf{E}_t(\alpha_{j,t+h})X_{j,t} - (1+r_0)^{-1}Y_t$ at the risk-free rate r_1 then after h periods one closes the positions.

The profit and its conditional expectation are:

$$\begin{aligned}
\mathbf{\Pi}_{t+h} &= Y_{t+h} - \textstyle\sum_{j=1}^{K}\mathsf{E}_t(\alpha_{j,t+h})X_{j,t+h} \\
&\quad +(1+r_1)\left\{\textstyle\sum_{j=1}^{K}(1+r_j)^{-1}\mathsf{E}_t(\alpha_{j,t+h})X_{j,t} - (1+r_0)^{-1}Y_t\right\} \\
\mathsf{E}_t(\mathbf{\Pi}_{t+h}) &= (1+r_1)\left\{\textstyle\sum_{j=1}^{K}(1+r_j)^{-1}\mathsf{E}_t(\alpha_{j,t+h})X_{j,t} - (1+r_0)^{-1}Y_t\right\} \\
&\quad > 0,
\end{aligned}$$

while the conditional variance of the profit is given by equation (3.3).

The mean-variance investment strategy is equivalent to the one proposed by Christoffersen and Giorgianni (2000) in that the two strategy imply exactly the same profit. Therefore, an investor is indifferent between the two strategies. Nevertheless, the derivation of the strategy by means of the mean-variance hedging approach highlights under which conditions speculation on

the interest rate differentials can be convenient, i.e. when inequality (3.4) is fulfilled. Furthermore, the mean variance approach also provides a simple mean of insuring a short position in a portfolio of the hard currencies when the initial capital is lower than the one required by perfect hedging. An explicit comparison with the hedging strategy derived by Christoffersen and Giorgianni (2000) is provided in the Appendix.

3.3 The estimation problem

This section discusses the problems connected with the estimation of the basket weights and the conditional variance of the profits. The estimates of the basket weights are needed in order to take an investment decision at a time t, but the investment outcome will be known at a future time $t + h$. Therefore, only observations up to time t can be used for the estimation. Furthermore, if the investment decision is made independently at every date: t, $t + 1$, ..., then one has to consider an "on-line" or "real-time" estimator which regularly updates the value of the estimate as a new observation becomes available. Finally, it is important to take into account the possible randomness of the basket weights, because the success of the investment strategy directly depends on the accuracy of the estimation.

3.3.1 Adaptive window estimation

An adaptive estimation procedure is now proposed which can cope with the problem of the on-line estimation of the time-varying regression coefficients of a system with stochastic regressors. The related statistical theory has been developed by Lipster and Spokoiny (1999), while an application to the estimation of the volatility of financial time series and the related theoretical properties can be found in Mercurio and Spokoiny (2004), in Härdle et al. (2000) and in Härdle et al. (2003). The regression coefficients are assumed to be locally time homogeneous, i.e. there exists some time interval where they can be well approximated by a constant. Consider the regression equation:

$$Y_t = \mathbf{X}_t^\top \alpha_t + \xi_t, \text{ with } \xi_t \sim \mathrm{N}(0, \sigma^2) \quad \text{for} \quad t = 1, \ldots, n, \tag{3.5}$$

where $\mathbf{X}_t$ and α_t are K-dimensional random vectors with finite second moments both independent from each other and from ξ_t. The assumption of local time homogeneity means that α_t is constant or at least nearly constant within an interval $I = [n-m, n]$ with $n - m > 0$, and n, $m \in \mathbb{N}$. Ideally, only the observations in the interval $I = [n-m, n]$ should be used for the estimation of α_n. Actually, an estimator of α_n using the observation of a subinterval $J \subset I$ would be less efficient, while an estimator using the observation of a larger interval $K \supset I$ would be biased. The main objective of the procedure which we propose is therefore to determine the largest interval of time homogeneity in a data driven way. Over this interval the estimation of the parameters can be carried out with ordinary least squares.

Some of the properties of the ordinary least squares (OLS) estimator

$$\widehat{\alpha}_I = W_I \sum_{t \in I} \mathbf{X}_t Y_t \tag{3.6}$$

are analyzed over the interval of time homogeneity $I = [n-m, n]$, where

$$W_I = \left(\sum_{t \in I} \mathbf{X}_t \mathbf{X}_t^\top \right)^{-1}. \tag{3.7}$$

Due to our assumption of local homogeneity, the value of α_t is close to a constant vector for each $t \in I$. This means that the estimation error:

$$\Delta_I = \sup_{t \in I} ||\alpha_t - \alpha_n||_2$$

is small. Denote the elements of the matrix W_I by $w_{ij,I}$, $i, j = 1, \ldots, K$. In the case of a standard regression model with deterministic design, the estimate $\widehat{\alpha}_I$ is the least squares estimate and $\sigma^2 W_I$ is its covariance matrix. In particular each diagonal element $w_{ii,I}\sigma^2$ of this matrix is the variance of the estimator $\widehat{\alpha}_{i,I}$, $i = 1, \ldots, K$. In this situation the design points are random. By analogy with the regression case, $w_{ii,I}\sigma^2$ is called the conditional variance of $\widehat{\alpha}_{i,I}$, $i = 1, \ldots, K$. Since the matrix $\sum_{t \in I} \mathbf{X}_t \mathbf{X}_t^\top$ is random, a random set is introduced where certain regularity conditions are satisfied. In particular one has to ensure that $\sum_{t \in I} \mathbf{X}_t \mathbf{X}_t^\top$ is invertible and in the sequel considerations are restricted to this set. For some positive constants $b > 0$, $B > 1$, $\rho < 1$, $r \geq 1$, $\lambda > \sqrt{2}$ and for $i = 1 \ldots, K$ define the random

set, where the following conditions are fulfilled:

$$A_{i,I} = \left\{ \begin{array}{c} b \leq w_{jj,I}^{-1} \leq bB; \\ \\ w_{jj,I} || \sum_{t \in I} \mathbf{X}_t \mathbf{X}_t^\top ||_\infty \leq r; \\ \\ |w_{ji,I}/w_{jj,I}| \leq \rho \quad \forall\, i = 1, \ldots, K \end{array} \right\},$$

where $|| \cdot ||_\infty$ denotes the supremum norm. Under these assumptions Härdle et al. (2000) have derived the following exponential probability bound:

$$\begin{aligned} &\mathsf{P}\left(|\widehat{\alpha}_{i,I} - \alpha_{i,n}| > \Delta_I + \lambda\sigma\sqrt{w_{ii,I}}\,; A_{i,I}\right) \\ &\quad \leq \quad 2\mathcal{P}(\lambda)\exp(-\lambda^2/2), \qquad i = 1, \ldots, K, \end{aligned} \tag{3.8}$$

where $\mathcal{P}(\lambda)$ is a polynomial in λ. It is remarked that the above result essentially generalizes the exponential bound for a regression model with Gaussian errors and deterministic design, where no regularity set is needed, $\Delta_I = 0$ and $\mathcal{P}(\lambda) = 1$.

The bound (3.8) can be used to estimate the coefficients α_t in the regression equation (3.5) when the regressors are random (for example lagged values of Y_t) and the coefficients are not constant, but locally time homogeneous.

Suppose that one expects time homogeneity in the interval I and hence in every subinterval J of I. This implies that the value Δ_I is negligible and similarly for all Δ_J, $J \subset I$ and that the mean values of the α_t over I and over J nearly coincide. Furthermore, it is known on the basis of equation (3.8) that the events

$$|\widehat{\alpha}_{i,I} - \alpha_n| \leq \mu\sigma\sqrt{w_{ii,I}} \quad \text{and} \quad |\widehat{\alpha}_{i,J} - \alpha_n| \leq \lambda\sigma\sqrt{w_{ii,J}}$$

occur with high probability for some sufficiently large constants λ and μ. By the triangle inequality, the following bound:

$$|\widehat{\alpha}_{i,I} - \widehat{\alpha}_{i,J}| \leq \mu\sigma\sqrt{w_{ii,I}} + \lambda\sigma\sqrt{w_{ii,J}}, \tag{3.9}$$

also holds under the assumption of homogeneity within I. Therefore, if an interval $J \subset I$ exists such that the inequality (3.9) is not fulfilled, then the hypothesis of homogeneity for the interval I is rejected. Finally, the adaptive estimator corresponds to the largest interval I such that the hypothesis of homogeneity is not rejected for I itself and all smaller intervals.

In practice we construct a multiple testing procedure. The time axis is divided in a regular grid with grid step $m_0 \in \mathbb{N}$. The right end point is equal n and the left end point is equal 1: i.e. the typical element of the grid is of the form:

$$t_k = \max\{1, \quad n - km_0\} \qquad \text{for} \qquad k = 0,\, 1,\, 2, \ldots .$$

Time homogeneity it is assumed for the smallest interval $[t_1,\, t_0]$ and it is tested within the next largest interval $[t_2,\, t_0]$ by comparing the OLS estimate related to it with the one related to $[t_2,\, t_1[$ and $[t_1,\, t_0]$ if they both fulfill the bound (3.9), then homogeneity is accepted for $[t_2,\, t_0]$. In this case, homogeneity is tested for the interval $[t_3,\, t_0]$, by comparing the estimate over it with the one over all its subintervals with left end point t_3 and with right end point t_0, i.e. $[t_3,\, t_k[$ and $[t_k,\, t_0]$ with $k = 1$ and 2. If inequality (3.9) is always fulfilled homogeneity is accepted for $[t_3,\, t_0]$ and the test is repeated for the next intervals $[t_k,\, t_0]$ in an analogous way until the hypothesis is not rejected. The interval of time homogeneity is estimated as the largest non rejected $\widehat{I} = [t_k,\, t_0]$. The adaptive estimator $\widehat{\alpha}_n$ of α_n is defined by applying the OLS over the selected interval $\widehat{I}$:

$$\widehat{\alpha}_{i,n} = \widehat{\alpha}_{i,\widehat{I}} \text{ for } i = 1, \ldots, K.$$

Notice that the previously described procedure requires the knowledge of the variance σ^2 of the errors ξ_t. In practical applications, σ^2 is typically unknown and has to be estimated from the data. The regression representation (3.5) and local time homogeneity suggests to apply a residual based estimator. Given an interval I the parameter estimate $\widehat{\alpha}_I$ is constructed. Next, the residuals $\widehat{\xi}_t$ are defined as $\widehat{\xi}_t = Y_t - \mathbf{X}_t^\top \widehat{\alpha}_I$. Finally, the variance estimator is defined by averaging the residuals squared:

$$\widehat{\sigma}^2 = |I|^{-1} \textstyle\sum_{t \in I} \widehat{\xi}_t^2.$$

3.3.2 The choice of m_0, λ and μ

The results of the procedure depends on the parameters m_0, λ and μ. If m_0 is small the grid is dense and the test of homogeneity is performed very often. On one hand, this increases the sensitivity of the procedure to structural changes, but on the other hand, it also increases the possibility of rejecting a large interval of time homogeneity. As far as change point models are concerned, the value of m_0 also implies a minimal delay in the perception of a jump. Both for simulation and real applications we choose $m_0 = 30$. Actually, repeated trials with different values have shown that moderate changes to the value of m_0 do not have great influence on the results.

The values of λ and μ influence the estimation procedure like the bandwidth in nonparametric regression (Green and Silverman, 1994). They determine the sensitivity of the adaptive estimation procedure. Smaller values of λ and μ are likely to lead to a rejection of large intervals, so that the estimate tends to have a smaller bias, but a larger variance. Larger values of λ and μ present the opposite problem because they reduce the sensitivity of the procedure to change points.

For the simulations we chose $\lambda = 2$ and $\mu = 4$, while for the real application we select the parameters in a data driven way. Attention is restricted to a small set of pairs: $\mathcal{S} = \{(\lambda, \mu)|\ \lambda, \mu \in \{2, 4, 6, 8\}\}$, and all the pairs are compared with the following criterion at each date t:

$$(\lambda^*, \mu^*) = \underset{(\lambda,\mu)\in\mathcal{S}}{\arg\min} \sum_{s=t-200}^{t-1} \left(Y_s - \sum_{j=1}^{K} \widehat{\alpha}_{j,s|s-h} X_{j,s} \right)^2 .$$

Finally, the value of $\widehat{\alpha}_{t+h|t}$ is estimated with the selected pair (λ^*, μ^*), i.e. the one which has provided the least quadratic hedging costs over the past trading periods.

3.3.3 Monte Carlo simulation

In order to illustrate the performance of the adaptive estimation procedure a small simulation study is made. The following process is considered:

$$Y_t = \alpha_{1,t} + \alpha_{2,t}X_{2,t} + \alpha_{3,t}X_{3,t} + \xi_t.$$

The length of the sample is 300. The regressors X_2 and X_3 are two independent random walks. The regressor coefficients are constant in the first half of the sample, then they make a jump after which they continue being constant until the end of the sample. For such a model the *interval of time homogeneity* coincides with the period where the regression coefficients are constant and the main issue is to detect the change point as quick as possible. We simulate three models with jumps of different magnitude. The values of the simulated models are presented in Table 3.1. The error term ξ

Table 3.1: Simulated models.

$1 \le t \le 150$		$151 \le t \le 300$	
Starting Values	Large Jump	Medium Jump	Small Jump
$\alpha_{1,t} = 1$	$\alpha_{1,t} = 1.005$	$\alpha_{1,t} = 1.0025$	$\alpha_{1,t} = 1.001$
$\alpha_{2,t} = 1$	$\alpha_{2,t} = 1.005$	$\alpha_{2,t} = 1.0025$	$\alpha_{2,t} = 1.001$
$\alpha_{3,t} = 1$	$\alpha_{3,t} = 1.005$	$\alpha_{3,t} = 1.0025$	$\alpha_{3,t} = 1.001$

is a Gaussian white noise, with zero mean and variance $\sigma = 0.001$. For each of the three models above 100 realizations of the white noise ξ are generated and the adaptive estimation is performed. Note that the average value of $\sigma|\xi_t|$ equals $10^{-2}\sqrt{2/\pi} \approx 0.008$, therefore even the large jump of magnitude 0.005 is generally not visible by eye. Figure 3.1 shows the true value of the coefficients along with the median, the 25% upper and lower quantile of the estimates of all realizations for each model at each time point. The plots of the bottom line display the median and the 25% upper and lower quantile of the on-line estimate of the length of the interval of time homogeneity. The

simulation results are very satisfactory. The change point is quickly detected, almost within the minimal delay of 30 periods for all three models, so that the adaptive estimation procedure show a good performance even for the small jump model. Note that the estimate of the intercept coefficient has quite a large variability, but this fact is only related to the slower rate of convergence with respect to the estimates of the other two coefficients and not to the proposed jump detection procedure.

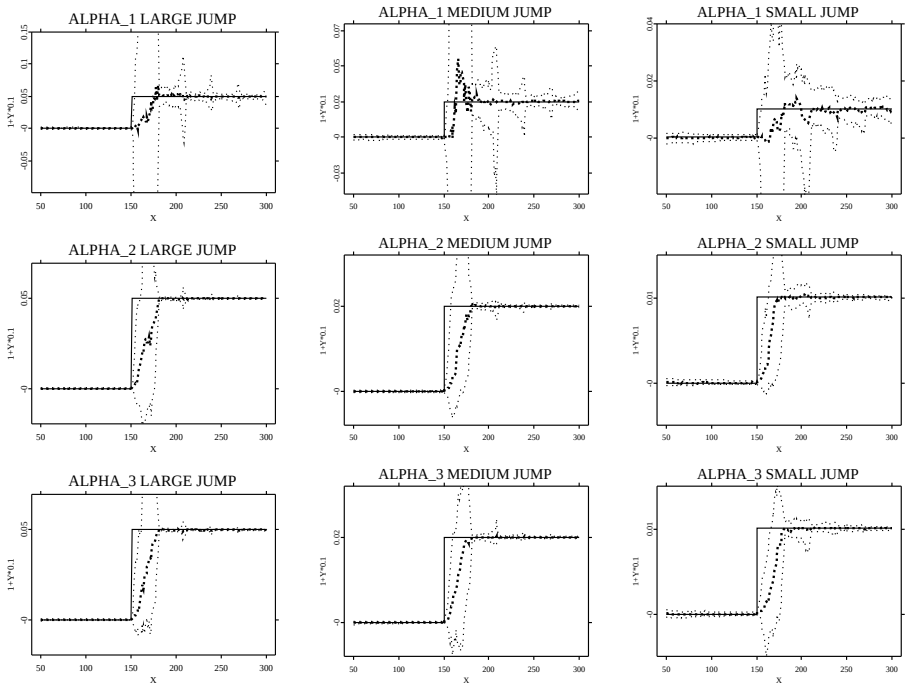

Figure 3.1: On-line estimates of the regression coefficients with jumps of different magnitude and of the intervals of time homogeneity. Median (thick dotted line), 25% lower and upper quantile (thin dotted line) among all estimates.

3.3.4 Three benchmark models

In order to assess the validity of the adaptive estimator its performance is compared with three procedures which are commonly used for on-line esti-

mation. All these methods are used in the next section for the empirical analysis of the Thai Baht Basket.

The most simple approach to the estimation of the basket weights is the recursive OLS. This algorithm suits the problem very well if the basket weights are indeed constant. In this case (3.1) can be seen as a regression equation, and one has only to perform an OLS estimation at each time t, with all the observations available at that time. The regressors are exchange rates which, since Meese and Rogoff (1983), have been usually modeled as random walks. The OLS estimator is therefore even super-consistent, that is it converges at rate t instead of $\sqrt{t}$, because we are estimating a cointegrated system (Hamilton, 1994).

But, if the basket weights are not constant, the recursive OLS produces in general very poor results. A pragmatic and popular way of taking into account the variability of the coefficients consists in choosing a rolling estimator. A fixed window of the most recent data is defined, where the estimation is performed. Once a new observation becomes available, the last observation is dropped from the end window and the new one is added at the beginning. Such an algorithm is very easy to implement. Nevertheless, it has many potential drawbacks. It deletes automatically from the sample many observations which could still fulfill the assumptions of parameter constancy, and it does not even try to prevent the fact that a structural break may be present just in the middle of the window which is currently used for the estimation. In this application a window of the last 250 observations is kept.

Another common approach to the estimation of time varying parameters is the Kalman filter, see e.g. Cooley and Prescott (1973), Granger (1986) and Canarella et al. (1990). In order to implement the Kalman filter one has to specify a Markov process for the parameters in order to describe explicitly their dynamics. Following Christoffersen and Giorgianni (2000), we specify the time varying basket weights α_t as a random walk:

$$\alpha_t = \alpha_{t-1} + \eta_t \qquad \eta_t \sim \mathrm{N}(0, \Sigma).$$

If the model is correctly specified and Σ and σ^2 are known, then α_t can be

estimated recursively with the Kalman filter, which has the property of being the best linear estimator. But the main drawback is that one neither knows σ^2 and Σ, nor if the model is correctly specified. For this reason one has to plug into the algorithm some "reasonable" values. This problem is actually very similar to choosing the two smoothing parameters in the nonparametric procedure, which we describe in Section 3.3.1. Similarly to Christoffersen and Giorgianni (2000) we estimate Σ and σ^2 on-line: σ^2 is estimated from the residuals of recursive OLS while Σ is estimated by taking the sample covariance matrix of the first differences of the past basket weights estimated with recursive OLS. The recursive OLS, the rolling OLS, the Kalman filter and the adaptive estimator require presample values in order to be initialized. For this purpose the first 350 observations are used which are then discarded.

3.3.5 The conditional variance of the profit

It is also important to compute the conditional variance of the profit in order to evaluate the risks of the investment. The estimation and in particular the forecast of the variance are quite difficult tasks, mainly because the realizations of the variance are not observed, so that one can hardly measure the goodness of the prediction. Furthermore in this case we have to provide an h-step ahead forecast, where $h = 30$ and 90, which is a very long horizon.

The expected variance of the profit, which is also the expected quadratic cost of hedging, is given by equation (3.3). Define now $\widehat{\alpha}_{j,t+h|t}$ as an estimator of $\mathsf{E}_t(\alpha_{j,t+h})$, and assume that the conditional variance of the profit is constant for all t. Then one can estimate it by averaging over the past realizations of the square hedging costs.

$$\hat{\boldsymbol{\sigma}}^2_{t+h|t} = (t-h)^{-1} \sum_{s=1}^{t-h} \left(Y_{s+h} - \sum_{j=1}^{K} \widehat{\alpha}_{j,s+h|s} X_{j,s+h} \right)^2 . \tag{3.10}$$

The assumption that the conditional variance is constant may sound quite restrictive. Nevertheless the above estimator appears to be sensible even in the presence of conditional heteroskedasticity, because we are interested in 30 and 90 step ahead forecast which are usually close to the unconditional variance. The above formulation of the estimator of the conditional variance

has the advantage of being model free and it has the appealing interpretation of directly linking the risk of choosing an estimator to its past performance.

3.4 An application to the Thai Baht basket

The purpose of this section is twofold: on one hand, it investigates whether arbitrage profits were possible among the currencies composing the Thai Baht basket, on the other it evaluates how the results of the speculation are affected by the choice of the estimator.

A similar analysis on the same data set has been performed by Christoffersen and Giorgianni (2000). As far as the comparison of the estimator is concerned, they argue that the Kalman filter should be preferred to the recursive and rolling estimators. This conclusion however is not supported by the results of this study.

The data set contains the daily exchange rates of the Thai Baht (THB), Japanese Yen (JPY) and German Mark (DEM) against the US Dollar (USD), together with the nominal inter-bank 1-and 3-months interest rates on THB, JPY, DEM and USD deposits. Figure 3.4 shows the exchange rates and Figure 3.3 the 3-months interest rates. The plot of the 1-months interest rates displays a similar pattern to the one of the 3-months interest rates and it is therefore omitted. The period under observation is January 2 1992 to February 12 1997. The source of the data is Bloomberg, L.P. and they were kindly provided by Lorenzo Giorgianni from IMF
.

From 1985 until its suspension on July 2, 1997 (following a speculative attack) the Baht was pegged to a basket of currencies consisting of Thailand's main trading partners. In order to gain greater discretion in setting monetary policy, the Bank of Thailand neither disclosed the currencies in the basket nor the weights. Similarly to Christoffersen and Giorgianni (2000) we assume to know the currencies composing the basket: USD, JPY and DEM. Therefore

the USD/THB exchange rate can be expressed in the following way:

$$Y_{USD/THB,t} = \alpha_{USD} + \alpha_{DEM} X_{USD/DEM,t} + \alpha_{JPY} X_{USD/JPY,t} + \xi_t.$$

The above equation seems to be confirmed by the statistical evidence because the R^2 with respect to the full sample estimation is around 0.8 and the estimated coefficients with fully modified OLS (Hamilton, 1994) are highly significant. Unit root tests confirm the hypothesis of nonstationarity for the univariate exchange rate time series, while Mean-F and Sup-F tests (Hansen, 1992) reject the hypothesis of a stable cointegration relationship among them (Christoffersen and Giorgianni, 2000).

3.4.1 The results

The estimation results can be seen in Figure 3.2 which show the output of the adaptive procedure calibrated for a 3-month forecast horizon together with the recursive and rolling OLS estimates. The results for the 1-month horizon looks similar and are therefore not displayed.

It is interesting to see that the adaptive estimate tends to coincide with the recursive estimate during the first half of the sample, more or less, while during the second half of the sample it tends to follow the rolling estimate. This may be seen as a hint of a change point, possibly a large realignment of the basket weights. However, these results may be also due to the large outlier which is visible in the middle of the THB/USD exchange rate sample (Figure 3.4).

The main criterion for evaluating the performance of the different estimation procedures is made by plugging the estimated values of the basket weights into the formula of the optimal quadratic risk minimizing strategy developed in Section 3.2. The results are displayed in Table 3.2. The investments are normalized such that at each trading day we take a short position of 100 USD in the optimal portfolio of the hard currencies. The result refers to the period April 9 1993 to February 12 1997 for the one month horizon investment and June 7 1993 to February 12 1997 for the three month horizon investment. The average expected profits, the average expected standard

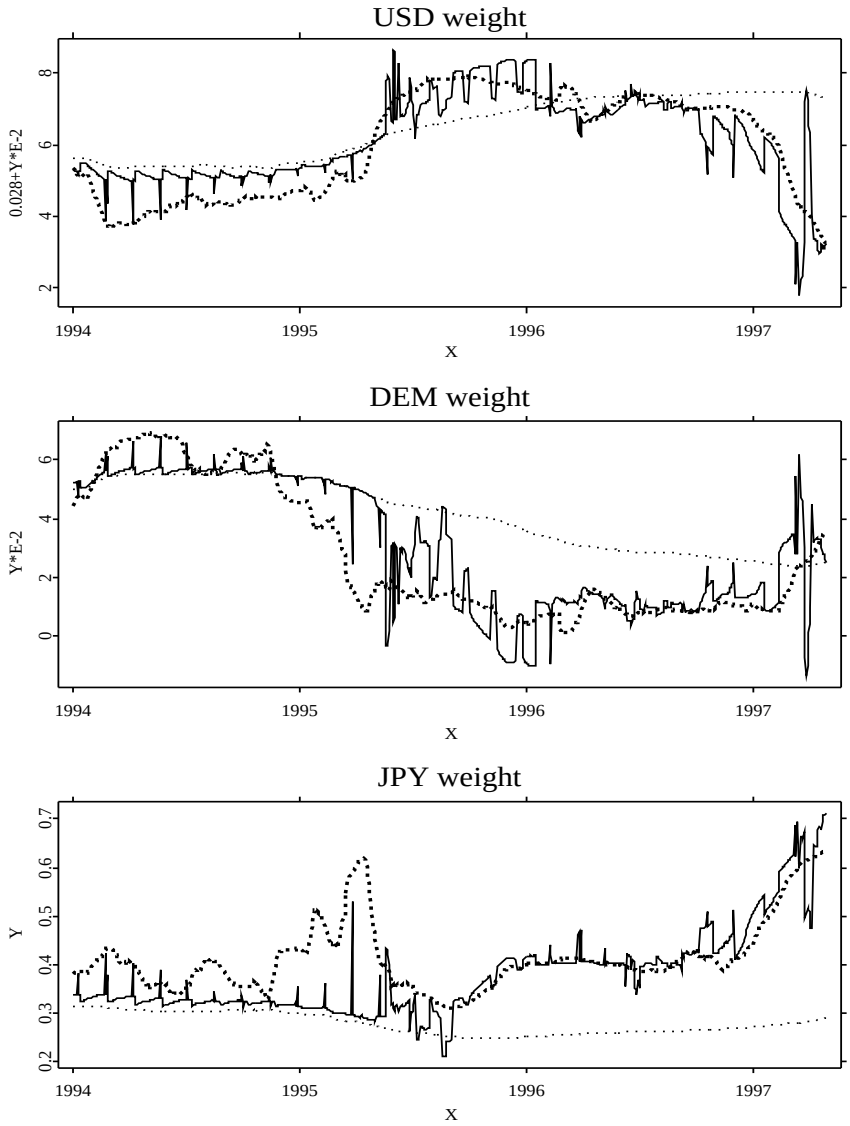

Figure 3.2: Estimated exchange rate basket weights: 3-month horizon adaptive (straight line), recursive (thin dotted line), rolling (thick dotted line).

deviations of the profit and the average realized profit are computed. The cumulative profits, i.e. the sum over all the realizations of the profits is also calculated. This quantity is quite interesting since it expresses how much

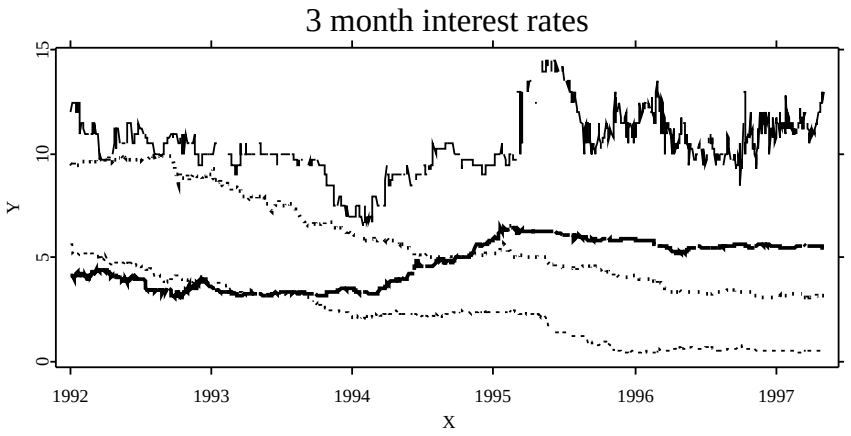

Figure 3.3: 3-month interest rate time series

one could have lost or gained by preferring a specific estimator.

The *expected profits* are on the average larger than the realized profits for all the estimators, so that the expectations are in general upward biased. The largest bias is due by far to the recursive estimator, while the Kalman filter shows the smallest one although the adaptive and rolling methods are quite close.

The *expected standard deviation* is similar again across the Kalman filter, rolling and adaptive estimator, while the recursive OLS shows larger results for both investment horizons. This hints at its poor forecasting performance which is probably due to the randomness of the basket weights.

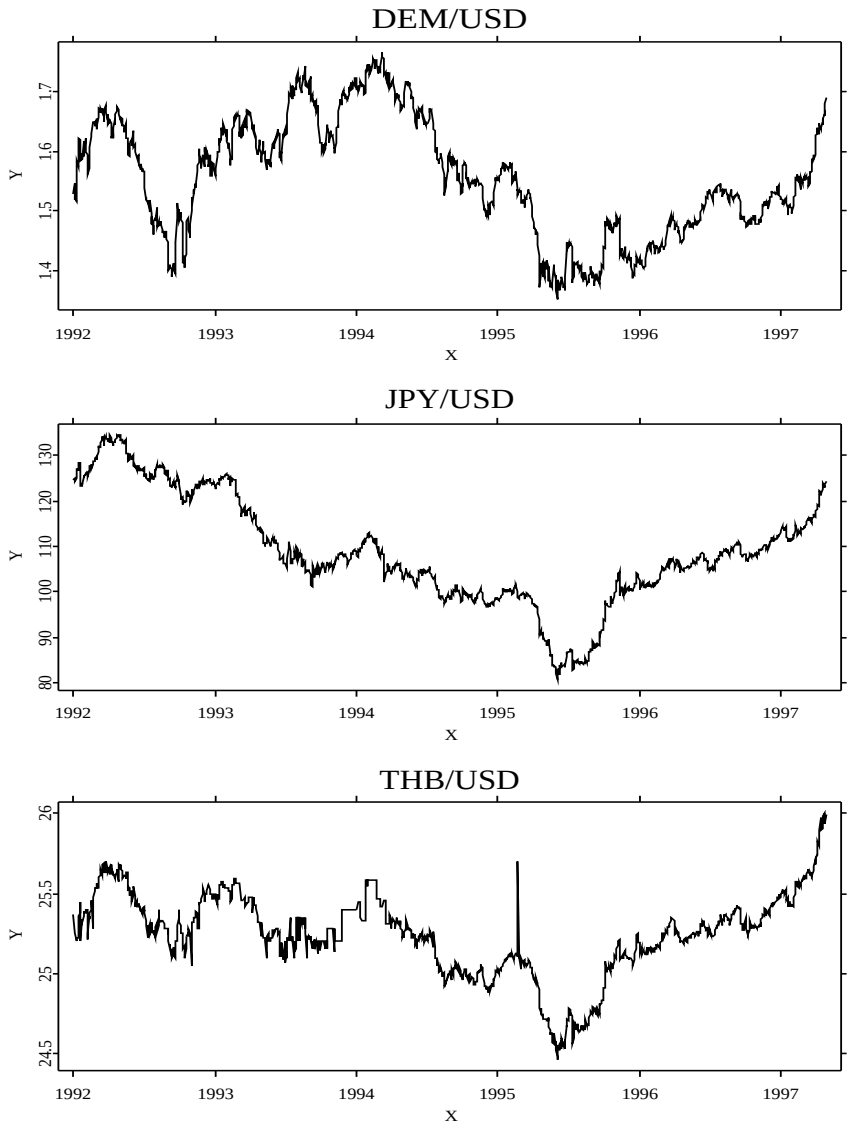

Figure 3.4: Exchange rates time series and interest rates time series: German (thick dotted line), Japanese (thin dotted line), American (thick straight line), Thai (thin straight line).

The *average realized profits* are quite similar among all methods. They are positive and, as far as the three month investment horizon is concerned, they are significantly larger than zero. This provides a clear evidence for the fact that arbitrage profits were possible with in the framework of the Thai Baht basket for the period under study. The adaptive estimator obtains the largest profits for both investment horizons.

Finally the importance of choosing a good statistical tool is confirmed by the *cumulative profits* which show that even a small improvement in the average profit can be quite rewarding eventually. The best performance is obtained by the adaptive estimator, while the second best performance is obtained by the recursive and rolling OLS together for the one month horizon and by the recursive OLS alone for the three month horizon.

Table 3.2: Summary statistics of the profits.

One Month Horizon	Recursive	Rolling	KF	Adaptive
Average Expected Profits	0.77	0.56	0.50	0.55
Average Expected Std. Deviation	0.54	0.35	0.33	0.39
Average Realized Profit	0.40	0.40	0.38	0.42
Standard Errors	0.30	0.30	0.33	0.33
Cumulative Profit	344.10	344.60	331.60	360.80
Three Month Horizon	**Recursive**	**Rolling**	**KF**	**Adaptive**
Average Expected Profits	1.62	1.46	1.37	1.45
Average Expected Std. Deviation	0.63	0.54	0.47	0.45
Average Realized Profit	1.16	1.14	1.14	1.18
Standard Errors	0.46	0.51	0.47	0.43
Cumulative Profit	945.40	925.70	929.10	958.50

3.5 Conclusions

The problem of investing and making arbitrage profits in a capital market where the currencies are linked to a basket has been analyzed. First, the issue of investing in an exchange rate basket has been considered from a finance theoretical perspective. In a mean-variance hedging setting, the optimal investment strategy, its expected profits and its conditional variance have been derived and the conditions for a profitable speculation on interest rate differentials have been highlighted.

Secondly, the statistical problems connected with the implementation of the arbitrage strategy and the numerical evaluation of the expected profits and conditional variances have been tackled. In particular, the solution to these problems rests on the estimation of the basket weights. A new adaptive estimation strategy has been proposed, which basically assumes that the coefficients are at least almost constant over some unknown intervals. The procedure determines this interval of time homogeneity, where the estimation can be carried out with ordinary least square. A small simulation study shows that the adaptive estimator performs quite well for change point models.

Finally, the above results have been tested in an empirical study of the Thai Baht exchange rate basket. Specifically, the optimal investment strategy has been implemented and the profits evaluated. The values of the basket weights have been estimated by means of the adaptive procedure. Three benchmark models have also been considered: the recursive OLS, the rolling OLS and the Kalman filter. The first one neglects the possible randomness in the parameters, the second one accounts for the time variability of the basket weights because it performs the estimation only on a moving window of the data of fixed length and the third assumes explicitly that the parameters follow a random walk. All the estimators provide positive arbitrage profits, however the results clearly show the importance of taking the possible time variability of the parameters into account, in particular for a correct evaluation of the risk. The performance of the adaptive method appears to be the

best one. The economic rationale for this result is the following: given that the Central Bank intervenes occasionally on the basket weights, the type of time variation in the basket weights is better described by a jump process and the adaptive estimator is intrinsically more suitable for the estimation of jump processes.

3.6 Appendix

This appendix shows the equivalence between the mean-variance hedging strategy and the one suggested by Christoffersen and Giorgianni (2000). They consider a long position in Y_{t+h} and a short position in a portfolio of the currencies which compose the basket: $\sum \vartheta_j X_{j,t+h}$. Their expression for the profit is therefore:

$$\widetilde{\mathbf{\Pi}}_{t+h} = Y_{t+h} - \sum_{j=1}^{K} \vartheta_j X_{j,t+h} = \sum_{j=1}^{K} \alpha_{j,t+h} X_{j,t+h} + \xi_{t+h} - \sum_{j=1}^{K} \vartheta_j X_{j,t+h}$$

and they impose the condition that the initial cash flow has to be zero:

$$0 = \sum_{j=1}^{K} (1+r_j)^{-1} \vartheta_j X_{j,t} - (1+r_0)^{-1} Y_t.$$

The investment strategy $(\vartheta_1, \ldots, \vartheta_K)$ is constructed in order to remove the currency $j = 2, \ldots, K$ risk (on the average) and therefore it must satisfy the following linear system of equations:

$$\begin{aligned} 0 &= \mathsf{E}_t \left(\partial \widetilde{\mathbf{\Pi}}_{t+h} / \partial X_{j,t+h} \right) = \mathsf{E}_t(\alpha_{j,t+h}) - \vartheta_j, \quad j = 2, \ldots, K \\ 0 &= \sum_{j=1}^{K} (1+r_j)^{-1} \vartheta_j X_{j,t} - (1+r_0)^{-1} Y_t. \end{aligned}$$

The solution is:

$$\begin{aligned} \vartheta_j &= \mathsf{E}_t(\alpha_{j,t+h}), \qquad j = 2, \ldots, K \\ \vartheta_1 &= (1+r_1) \left\{ (1+r_0)^{-1} Y_t - \sum_{j=2}^{K} (1+r_j)^{-1} \mathsf{E}_t(\alpha_{j,t+h}) X_{j,t} \right\}. \end{aligned}$$

the investment strategy, based on mean-variance hedging arguments, is equivalent to the one of Christoffersen and Giorgianni (2000) because the profits are equal: $\mathbf{\Pi}_{t+h} = \widetilde{\mathbf{\Pi}}_{t+h}$. Recall that $X_{1,t} \equiv 1$, than it is easy to see that:

$$\begin{aligned} \mathbf{\Pi}_{t+h} &= Y_{t+h} - \sum_{j=1}^{K} \psi_j^* X_{j,t+h} + (1+r_1) \left\{ \sum_{j=1}^{K} \psi_j^* X_{j,t} (1+r_j)^{-1} - Y_t (1+r_0)^{-1} \right\} \\ &= Y_{t+h} - \sum_{j=2}^{K} \psi_j^* X_{j,t+h} + (1+r_1) \left\{ \sum_{j=2}^{K} \psi_j^* X_{j,t} (1+r_j)^{-1} - Y_t (1+r_0)^{-1} \right\} \\ &= Y_{t+h} - \sum_{j=1}^{K} \vartheta_j X_{j,t+h} = \widetilde{\mathbf{\Pi}}_{t+h}. \end{aligned}$$

Chapter 4

Estimation of time dependent volatility via local change point analysis

4.1 Introduction

Since the seminal papers of Engle (1982) and Bollerslev (1986), modeling the dynamic features of the variance of financial time series has become one of the most active fields of research in econometrics. New models, different applications and extensions have been proposed as it can be seen by consulting, for example, the monographs of Engle (1995) and of Gouriéroux (1997). The main idea behind this strain of research is that the volatility clustering effect that is displayed by stock or exchange rate returns can be modeled globally by a stationary process. This approach is somehow restrictive and it does not fit some characteristics of the data, in particular the fact that the volatility process appears to be "almost integrated" as it can be seen by usual estimation results and by the very slow decay of the autocorrelations of squared returns. Other global parametric approaches have been proposed by Engle and Bollerslev (1986) and by Baillie et al. (1996) in order to include these features in the model. Furthermore, continuous time models, and in particular diffusions and jump diffusions, have also been considered; see for example Andersen et al. (2002) and Duffie et al. (2000).

However, Mikosch and Starica (2000b) showed that long memory effects of financial time series can be artificially generated by structural breaks in the parameters. This motivates another modeling approach which borrows its philosophy mainly from the nonparametric statistics. The main idea consists in describing the volatility clustering effect only by a locally stationary process. Therefore, only the most recent data are considered and weighting schemes, which can be themselves either global or local and data driven, are suggested in order to decrease the dependence of the estimate on the older observations. Some examples of this approach can be found in Fan and Gu (2003), in Dahlhaus and Rao (2003) and in Cheng et al. (2003). Furthermore, Mercurio and Spokoiny (2004) (referred to as MS2004 in what follows) propose a new local adaptive volatility estimation (LAVE) of the unknown volatility from the conditionally heteroskedastic returns. The method is based on pointwise data-driven selection of the interval of homogeneity for every time point. The numerical results demonstrate a reasonable performance of the new method. In particular, it slightly outperforms the standard GARCH(1,1) approach. Härdle et al. (2003) extend this method to estimating the volatility matrix of the multiple returns and Mercurio and Torricelli (2003) apply the same idea in the context of a regression problem.

The aim of this chapter is to develop another procedure which, however, applies a similar idea of pointwise adaptive choice of the interval of homogeneity. The main difference between the LAVE approach from MS2004 and the new procedure is in the way of testing the homogeneity of the interval candidate and in the definition of the selected interval. We systematically apply the approach based on the local change point analysis. This means that every interval is tested on homogeneity against a change point alternative. If the hypothesis is not rejected, a larger interval candidate is taken. If the change point is detected, then the location of the change point is used for defining the adaptive interval while MS2004 suggested to take the latest nonrejected interval. The modified procedure allows to improve the sensitivity of the method to changes of volatility by using the more powerful likelihood ratio test statistic with the careful choice of the critical level. In addition,

the use of the additional information about the location of the change point which is delivered by the change point test, helps to reduce the estimation bias. Finally, the interpretation of the procedure as a multiple test against a change point alternative leads to a very natural method of tuning the parameters of the procedure.

The change point detection problem for financial time series was considered in Mikosch and Starica (2000a) but they focused on asymptotical properties of the test if only one change point is present. Kitagawa (1987) applied non-Gaussian random walk modeling with heavy tails as the prior for the piecewise constant mean for one-step-ahead prediction of nonstationary time series. However, the mentioned modeling approaches require some essential amount of prior information about the frequency of change points and their size. The new approach proposed in this chapter does not assume smooth or piecewise constant structure of the underlying process and does not require any prior information. The procedure proposed below in Section 4.3 focuses the on adaptive choice of the interval of homogeneity that allows to proceed in a unified way with smoothly varying coefficient models and change point models.

Another important feature of the proposed procedure is that it can be easily extended to multiple volatility modeling, cf. Härdle et al. (2003). Suppose that a number of financial time series is observed and the goal is to estimate the corresponding time depending volatility matrix. We again assume that the volatility matrix is nearly constant within some historical time interval which we identify from the data. The volatility matrix is estimated in a usual way from the observations which belong to the detected interval.

The chapter is organized as follows. The next two sections introduce the adaptive modeling procedure. Section 4.4 presents an extension to the multiple volatility case. Some theoretical properties of the procedure are discussed in the general situation and for two particular cases: a change point model with piecewise constant volatility and the case of a volatility function smoothly varying in time. Section 4.6 illustrates the performances of the

new methodology by means of some simulated examples and applications to real data sets. First we address the problem of selecting the smoothing parameters and propose one solution which is systematically applied for all the examples. Section 4.6.2 presents some numerical results for a change point model. In Section 4.6.4 we study forecasting ability of the new method by mean of a comparative study with the GARCH(1,1) method. Sections 4.6.7 and 4.6.8 discuss applications of the new method to the Value at Risk problem. Finally, Section 4.7 collects the proofs of the main results.

4.2 Volatility modeling. Univariate case

Let S_t be an observed asset process in discrete time, $t = 1, 2, \ldots$, while R_t defines the corresponding return process: $R_t = \ln(S_t/S_{t-1})$. We model this process via the *conditional heteroskedasticity* assumption:

$$R_t = \sigma_t \xi_t \,, \tag{4.1}$$

where ξ_t, $t \geq 1$, is a sequence of independent standard Gaussian random variables and σ_t is the *volatility* process which is in general a predictable random process, that is, σ_t is measurable with respect to $\mathcal{F}_{t-1}$ with $\mathcal{F}_{t-1} = \sigma(R_1, \ldots, R_{t-1})$ (σ-field generated by the first $t-1$ observations).

Similarly to MS2004 we focus on the problem of filtering the parameter σ_t from the past observations $R_1, \ldots, R_{t-1}$. This problem naturally arises as an important building block for many tasks of financial engineering like Value at Risk or Portfolio Optimization.

4.2.1 Parametric modeling

A *time-homogeneous* (*time-homoskedastic*) model means that σ_t is a constant. The process S_t is then a geometric Brownian motion observed at discrete time moments. For the homogeneous model $R_t = \sigma \xi_t$ with $t \in I$, the

parameter $\theta = \sigma^2$ can be estimated using the maximum likelihood method:

$$\tilde{\theta}_I = \operatorname*{argmax}_{\theta \geq 0} L_I(\theta) = \operatorname*{argmax}_{\theta \geq 0} \sum_{t \in I} \ell(R_t, \theta)$$

where $\ell(y, \theta) = -(1/2)\ln(2\pi\theta) - y^2/(2\theta)$ is the log-density of the normal distribution with the parameters $(0, \theta)$. A simple algebra yields

$$\tilde{\theta}_I = N_I^{-1} \sum_{t \in I} R_t^2, \qquad \text{and} \qquad L_I(\tilde{\theta}_I) = -\frac{N_I}{2} \ln(2\pi\tilde{\theta}_I) - \frac{N_I}{2} \tag{4.2}$$

where N_I denotes the number of time points in I.

The assumption of normality for the innovations ξ_t is often criticized in the financial literature. Our empirical examples in Section 4.6.4 below also indicate that the tails of estimated innovations are heavier than the normality would imply. However, the estimate $\tilde{\theta}_I$ remains meaningful even for the non-normal innovations, it is just a quasi-likelihood approach. Moreover, one can show that this approach leads to the same asymptotic quality of estimation if the distribution of the ξ_t's fulfills some exponential moment conditions.

4.3 Adaptive volatility estimation. Univariate case

The assumption of time homogeneity is too restrictive in practical applications and it does not allow to fit well real data. We consider an approach based on the *local time-homogeneity* which means that for every time moment n there exists a historic time interval $[n-m, n[$ in which the volatility process σ_t is nearly constant. Under such a modeling, the main intention is both to describe the interval of homogeneity and to estimate the corresponding value σ_n.

4.3.1 Choice of the interval of homogeneity by local change point analysis

Our approach is based on the adaptive choice of the interval of homogeneity for the end point n. This choice is made by the *local change point detection* (LCPD) algorithm described below. The procedure attempts to find this interval from the data by successive testing the hypothesis of homogeneity. We start by defining a family $\mathcal{I} = \{I_k, k = 0, 1, \ldots\}$ of intervals of the form $I_k = [n - m_k, n[$, where the index parameter m_k fulfills: $m_0 < m_1 < m_2 < \ldots \leq n$. The intervals I_k are naturally ordered by their length m_k. We begin from the smallest interval I_0 and we test the hypothesis of homogeneity within I_0 against a change point alternative. If the hypothesis is not rejected then we take the next larger interval and continue this way until we detect a change point or the largest possible interval $[0, n[$ is reached. If the hypothesis of homogeneity within some interval I_k is rejected, the estimate of the change point $\widehat{\nu}$ is computed within this interval and the estimated interval of homogeneity is defined as $\widehat{I} = [\widehat{\nu}, n[$, otherwise we take $\widehat{I} = [0, n[$. Finally, we estimate the volatility parameter $\theta_n = \sigma_n^2$ from the observations R_t for $t \in \widehat{I}$ assuming the homogeneous model within $\widehat{I}$, that is, we define $\widehat{\theta}_n = \tilde{\theta}_{\widehat{I}}$.

The main ingredient of this procedure is the homogeneity test for every interval I_k, which is described in the next section.

4.3.2 Test of homogeneity against a change point alternative

Let I be an interval candidate. The null hypothesis for $I = [n - m, n[$ means that the observations R_t for $t \in I$ follow the parametric model with the parameter θ. This hypothesis leads to the log-likelihood $L_I(\theta)$. We want to test this hypothesis against a change point alternative that the parameter θ spontaneously changes in some internal point τ of the interval I. Let $\mathcal{T}_I$ be a set of internal points within I. Every point $\tau \in \mathcal{T}_I$ splits the interval

I onto two subintervals $J = [\tau, n[$ and $J^c = I \setminus J = [n-m, \tau[$. The change point alternative means that $\theta_t = \theta$ for $t \in J$ and $\theta_t = \theta'$ for $t \in J^c$ for some $\theta \neq \theta'$. This corresponds to the log-likelihood $L_J(\theta) + L_{J^c}(\theta')$. The likelihood ratio test statistic for the change point alternative with the change point location at the point τ is of the form

$$\begin{aligned} T_{I,\tau} &= \max_{\theta,\theta'} \{L_J(\theta) + L_{J^c}(\theta')\} - \max_{\theta} L_I(\theta) \\ &= L_J(\tilde{\theta}_J) + L_{J^c}(\tilde{\theta}_{J^c}) - L_I(\tilde{\theta}_I) = \widehat{L}_J + \widehat{L}_{J^c} - \widehat{L}_I. \end{aligned}$$

For the considered volatility model, this test statistic can be represented in the form

$$T_{I,\tau} = N_J K(\tilde{\theta}_J, \tilde{\theta}_I) + N_{J^c} K(\tilde{\theta}_{J^c}, \tilde{\theta}_I)$$

where $K(\theta, \theta') = -0.5\,(\ln(\theta/\theta') - 1 + \theta/\theta')$ is the Kullback-Leibler information for the two normal distributions with variances θ and θ'. The change point test for the interval I is defined as the maximum of such defined test statistics over $\tau \in \mathcal{T}_I$:

$$T_I = \max_{\tau \in \mathcal{T}_I} T_{I,\tau}\,.$$

The change point test compares this statistic with the critical value λ_I which may depend on the interval I and the nominal first kind error probability α. The hypothesis of homogeneity is rejected if $T_I \geq \lambda_I$, in this case the estimator of the change point is defined as $\widehat{\nu} = \operatorname{argmax}_{\tau \in \mathcal{T}_I} T_{I,\tau}$. The way of choosing the critical value as well as the other parameters of the procedure like the set of testing intervals $\mathcal{T}_I$ is discussed in Section 4.3.3.

4.3.3 Parameters of the LCPD procedure

To start the procedure running, one has to specify some parameters. This includes the set $\mathcal{I}$ of interval candidates, and, for every $I \in \mathcal{I}$, the set of internal points $\mathcal{T}_I$ and the critical value λ_I. First we briefly discuss how the sets $\mathcal{I}$ and $\mathcal{T}_I$ can be selected. Then we focus on the choice of the critical

values λ_I.

It is useful to take the set $\mathcal{I}$ of interval candidates in the form of an arithmetic or geometric grid. In both cases one has to fix the starting interval length m_0, that is, the first considered interval is of the form $[n-m_0, n[$. At every iteration this length is increased by adding resp. by multiplying with some fixed step. In our theoretical study we assume the maximal possible set of all intervals I with the length not less than m_0. This is a special case of an arithmetic grid with the step one. However, in order to reduce the computational burden, a geometric grid is implemented in the simulation study and in the applications to real data; this means that the length m_k of the interval I_k is defined as $m_k = [m_0 c^k]$ for some $c > 1$.

For every interval $I \in \mathcal{I}$, $I = [n-m, n[$, we define $\mathcal{T}_I$ as the set of all internal points of I separated away from the end-point. More precisely, for a fixed $\rho \le 1/3$, set $\mathcal{T}_I = \{t : n - m + \rho m \le t \le n - \rho m\}$. A reasonable choice for ρ is $\rho = 1/3$. The idea behind this choice is that the behavior of the log-likelihood test statistic $T_{I,\tau}$ becomes quite irregular when τ approaches the end-points of the interval I. Note also that for the points close to n, the test on change point has been already made on the earlier steps of the algorithm while for the points close to the left end point $n-m$, a test will be made at the next iterations. Our simulations results indicate that the procedure is quite stable w.r.t. the choice of the parameters like ρ and c.

On the contrary, the choice of the critical values λ_I is rather important. Larger values λ_I improve stability of the method under homogeneity but result in a low sensitivity to parameter changes while too small critical values lead to a large "false alarm" probability. The standard approach to choosing the critical values is to provide a prescribed first kind error probability, that is, in the homogeneous case, the "false alarm" probability should not exceed the given level α. Here we describe different possibilities for this choice.

We consider a homogeneous model $R_t = \sigma\xi_t$ with the constant volatility $\theta = \sigma^2$ and standard Gaussian innovations ξ_t. It is worth noting that the

particular value θ has no influence on the behavior of the procedure (it is cancelled in the expression for the test statistic T_I) and therefore we can assume $\theta = 1$. Thus, the probability model is completely specified and its properties can be evaluated by the Monte Carlo simulation.

Define for every I a value β_I in a way that $\sum_{I\in\mathcal{I}} \beta_I = \alpha$. A reasonable proposal is

$$\beta_I = \alpha N_I^{-1} \left(\sum_{I'\in\mathcal{I}} N_{I'}^{-1} \right)^{-1}$$

We also denote $\alpha_I = \sum_{I'\in\mathcal{I}(I)} \beta_{I'}$ where $\mathcal{I}(I) = \{I' : I' \in \mathcal{I}, I' \subseteq I\}$.

Now we tune the parameters of the procedure for the time homogeneous data generating process. The values λ_I are computed sequentially starting from the smallest interval I_0. For this interval we define λ_{I_0} as $(1-\alpha_{I_0})$-quantile of the test statistic T_{I_0}. Now, if the values $\lambda_{I_{k'}}$ have been already set for all $k' < k$, then we compute λ_{I_k} in such a way that the interval I_k (and thus, all the smaller intervals) is accepted with the frequency $1-\alpha_{I_k}$. This corresponds to the choice of the λ_{I_k}'s from the condition

$$\mathsf{P}\Big(\max_{k'\le k} T_{I_{k'}} / \lambda_{I_{k'}} > 1 \Big) = \alpha_{I_k}.$$

It is easy to check that such a choice is possible for any growing sequence of critical values α_{I_k}. However, the method is computationally intensive. Several proposals to simplify this choice are discussed below. One is based on the result of Theorem 4.3 from Section 4.5 that suggests to apply a critical value λ_I that grows linearly with $\ln(N_I)$, that is, $\lambda_I = a + b\ln(N_I)$. The constants a and b might (and should) depend on the nominal error level α and on the choice of the set $\mathcal{I}$ and $\mathcal{T}_I$ or, more specifically, on the parameters ρ and c. So, the following method can be recommended: for fixed values α, ρ and c, compute critical values λ_I for a few intervals I and adjust a linear relationship $\lambda_I = a + b\ln(N_I)$. We continue this discussion in Section 4.6 where an implementation of the procedure will be discussed in details.

4.4 Extension to multiple volatility modeling

In this section we discuss how the approach can be extended to modeling multiple time series. Let $S_t \in I\!R^d$ be a vector of observed asset processes in discrete time, $t = 1, 2, \ldots$ and R_t is the vector of the corresponding returns: $R_{t,m} = \ln(S_{t,m}/S_{t-1,m})\,,\ m = 1, \ldots, d\,.$ The *conditional heteroskedasticity* assumption reads in this case as

$$R_t = \Sigma_t^{1/2}\xi_t\,, \tag{4.3}$$

where $\xi_t\,,\ t \geq 1\,,$ is a sequence of independent standard Gaussian random vectors and Σ_t is a symmetric $d \times d$ *volatility* matrix which is in general a predictable random process, that is, Σ_t is measurable with respect to $\mathcal{F}_{t-1}$ with $\mathcal{F}_{t-1} = \sigma(R_1, \ldots, R_{t-1})$ (σ-field generated by the first $t-1$ observations).

A *time-homogeneous* (*time-homoskedastic*) model means that the matrix Σ_t is a constant. For the homogeneous model $R_t = \Sigma^{1/2}\xi_t$ with $t \in I\,,$ the parameter Σ can be estimated from the observations $R_t\,,\ t \in I\,,$ using the maximum likelihood method:

$$\tilde{\Sigma}_I = \operatorname*{argmax}_{\Sigma} L_I(\Sigma) = \operatorname*{argmax}_{\Sigma} \sum_{t \in I} \ell(R_t, \Sigma)$$

where $\ell(y, \Sigma) = -(1/2)\ln(\mathsf{det}\Sigma) - y^\top \Sigma^{-1} y/2$ for $y \in I\!R^d$ corresponds to the log-density of the normal distribution with the parameters $(0, \Sigma)\,.$ (We skip the unimportant constant term $-(d/2)\ln(2\pi)$ in the expression of the density.) It is well known that the solution of this problem in the class of all symmetric nonnegative matrices is given by the empirical covariance matrix $\tilde{\Sigma}_I$:

$$\begin{aligned} \tilde{\Sigma}_I &= N_I^{-1} \sum_{t \in I} R_t R_t^\top\,, \\ L_I(\tilde{\Sigma}_I) &= -\frac{N_I}{2}\ln(\mathsf{det}\tilde{\Sigma}_I) - \frac{1}{2}\sum_{t \in I} R_t^\top \tilde{\Sigma}_I^{-1} R_t \end{aligned} \tag{4.4}$$

where N_I denotes the number of time points in $I\,.$ Note that

$$\sum_{t \in I} R_t^\top \tilde{\Sigma}_I^{-1} R_t = \sum_{t \in I} \mathsf{tr}\left(\tilde{\Sigma}_I^{-1} R_t R_t^\top\right) = N_I \mathsf{tr}\left(\tilde{\Sigma}_I^{-1}\tilde{\Sigma}_I\right) = N_I d$$

so that

$$\widehat{L}_I = L_I(\tilde{\Sigma}_I) = -\frac{N_I}{2}\ln(\mathsf{det}\tilde{\Sigma}_I) - \frac{N_I d}{2}\,. \tag{4.5}$$

For estimation of the time-dependent volatility matrix we apply the same approach as in the univariate case. Namely, we estimate the volatility matrix Σ_n at a point n using the local MLE $\tilde{\Sigma}_{\widehat{I}}$ where $\widehat{I}$ is the historical interval to the last detected change point. The procedure reads exactly as in the univariate case, see Section 4.3.

Let I be an interval candidate. Here we describe a change point test within I based on the likelihood ratio test statistics.

The null hypothesis for $I = [n-m, n[$ means that the observations R_t for $t \in I$ follow the parametric model with the volatility matrix Σ. This hypothesis leads to the log-likelihood $L_I(\Sigma)$. We want to test this hypothesis against a change point alternative that the volatility matrix Σ spontaneously changes in some internal point τ of the interval I. Similarly to the univariate case, the likelihood ratio test statistic for the change point alternative with the change point location at the point τ is of the form

$$\begin{aligned} T_{I,\tau} &= \max_{\Sigma,\Sigma'}\{L_J(\Sigma) + L_{J^c}(\Sigma')\} - \max_{\Sigma} L_I(\Sigma) \\ &= L_J(\tilde{\Sigma}_J) + L_{J^c}(\tilde{\Sigma}_{J^c}) - L_I(\tilde{\Sigma}_I) = \widehat{L}_J + \widehat{L}_{J^c} - \widehat{L}_I. \end{aligned}$$

For the considered volatility model, due to (4.5), this test statistic can be represented as

$$T_{I,\tau} = N_J K(\tilde{\Sigma}_J, \tilde{\Sigma}_I) + N_{J^c} K(\tilde{\Sigma}_{J^c}, \tilde{\Sigma}_I)$$

where $K(\Sigma, \Sigma_0) = 0.5\left(\mathsf{tr}(\Sigma{\Sigma_0}^{-1} - I_d) - \ln(\mathsf{det}\Sigma/\mathsf{det}\Sigma_0)\right)$ is the Kullback-Leibler information for the two normal distributions with variances Σ and Σ_0. The change point test for the interval I is defined as the maximum of such defined test statistics over $\tau \in \mathcal{T}_I$:

$$T_I = \max_{\tau\in\mathcal{T}_I} T_{I,\tau}\,.$$

The hypothesis of homogeneity is rejected if $T_I \geq \lambda_I$, in this case the estimator of the change point is defined as $\widehat{\nu} = \operatorname{argmax}_{\tau \in \mathcal{T}_I} T_{I,\tau}$. The critical value λ_I may depend on the interval I, on the nominal first kind error probability α, and on the dimension of the vector R_t. The parameters of the procedure like λ_I and the set of testing intervals $\mathcal{T}_I$ can be selected similarly to the univariate case, see Section 4.3.3.

4.5 Theoretical properties

This section discusses some useful theoretical properties of the adaptively selected interval of homogeneity $\widehat{I}$ and then of the adaptive volatility estimate $\widehat{\theta}$ that corresponds to the selected interval $\widehat{I}$, that is, $\widehat{\theta} = \tilde{\theta}_{\widehat{I}}$. We state the most of our results for the case of one time series. An extension of the results to the case of multiple time series will be briefly discussed in Section 4.5.7.

We start by discussing the "false alarm" probability, that is, the probability that a good interval in which the hypothesis of homogeneity is nearly fulfilled is rejected by the change point test. We show that if the critical values λ_I are properly selected the procedure ensures a prescribed false alarm probability level. The standard way for proving such a result is based on the asymptotic expansion of the log-likelihood process $L(\theta)$. Here we briefly discuss this approach and then switch to the nonasymptotic one.

4.5.1 Asymptotic properties of the change point test under the null

The LR test statistic T_I introduced in the previous section has nice asymptotic properties. In particular, it weakly converges under the null hypothesis to some nondegenerated distribution. Moreover, this distribution is parameter free. The idea behind the mentioned asymptotic result is that in the homogeneous situation the likelihood ratio process can be approximated by some fixed transformation of the standard Wiener process. More precisely,

define for every $0 \le \tau_1 < \tau_2 \le 1$ the interval $J = [\tau_1 n, \tau_2 n]$. Then the likelihood ratio $L_J(\theta, \theta_0)$ based on the observations from this interval for $\theta = \theta_0(1 + un^{-1/2})$ can be (strongly) approximated by the random variable $(\tau_2 - \tau_1)^{-1} \{u(W_{\tau_2} - W_{\tau_1}) - u^2/2\}$ where W_τ is a standard Wiener process. This yields the approximation of the distribution of the statistic $2T_I$ by the maximum of $\tau^{-1}W_\tau^2 + (1-\tau)^{-1}(W_1 - W_\tau)^2 - W_1^2$ over $\tau \in [\rho, 1-\rho]$. We do not discuss this result in more details because its applicability is restricted to the case of a large interval I while the procedure starts with a small interval I_0 even if the sample size is large.

Therefore, we need a version of this result which applies to an arbitrary sample size. We present such nonasymptotic results for two cases: one for pure homogeneous situation with a constant volatility and the other one for a nearly homogeneous case.

4.5.2 "False alarm" probability under the null

Suppose that an interval I is fixed. Within the LCPD procedure we test for every such interval the "null" hypothesis that the process R_t is homogeneous within I. Our first result presents a bound for the deviation probability for the test statistic T_I which is used in the algorithm.

Theorem 4.1 *Let $I \in \mathcal{I}$ and $\theta_t = \theta_0$ for all $t \in I$. Then it holds for every $z \ge 0$*

$$\mathsf{P}\left(T_I \ge 2z\right) \le 4N_I e^{-z}.$$

In particular, for every $\beta \in (0,1)$, with $z = \ln(4N_I/\beta)$, it holds

$$\mathsf{P}\left\{T_I \ge 2\ln(4N_I/\beta)\right\} \le \beta.$$

The next result describes the probability of rejecting a homogeneous interval by our procedure. This means that every of the test statistics $T_{I'}$ for every smaller interval $I' \in \mathcal{I}(I)$ does not exceed the corresponding critical

value $\lambda_{I'}$. This is a multiple testing problem requiring a correction of the critical value for using the multiple test. In our theoretical study we apply the Bonferonni method: for every interval I we assign a first kind error probability β_I such that the sum of the β_I's does not exceed the prescribed value α. It is well known that the Bonferonni method is a bit conservative. Therefore, the result we give only an upper bound for the 'false alarm' probability.

In what follows we suppose that some values β_I, $I \in \mathcal{I}$, are fixed such that

$$\alpha_I \stackrel{\mathrm{def}}{=} \sum_{I' \in \mathcal{I}(I)} \beta_{I'} \leq \alpha.$$

The next result is a straightforward corollary of Theorem 4.1.

Theorem 4.2 *If $\theta_t \equiv \theta_0$ for all $t \in I$ and if $\lambda_I \geq 2\ln(4N_I/\beta_I)$ for all intervals $I \in \mathcal{I}$, then for every $I \in \mathcal{I}$*

$$\mathsf{P}\left(I \textit{ is rejected}\right) \leq \alpha_I\,.$$

Proof It suffices to only mention that by Theorem 4.1

$$\mathsf{P}\left(I \text{ is rejected}\right) \leq \sum_{I' \in \mathcal{I}(I)} \mathsf{P}\left\{T_{I'} \geq 2\ln(4N_{I'}/\beta_{I'})\right\}.$$

4.5.3 "False alarm" probability in the nearly homogeneous case

Here we consider a more general situation when the volatility coefficient θ_t can be approximated within I by a value θ_0 which is measurable w.r.t. $\mathcal{F}_{n-m-1}$. The violation from the homogeneity within I can be naturally measured by the values Δ_I^* and Δ_I defined by the equations

$$\Delta_I^* = \sup_{t \in I} |\theta_t/\theta_0 - 1| \qquad \text{and} \qquad \Delta_I^2 = N_I^{-1} \sum_{t \in I} |\theta_t/\theta_0 - 1|^2. \tag{4.6}$$

Note that in general Δ_I and Δ_I^* are random variables. Near homogeneity within I means that these values are small with a high probability. For every numbers $\mu \geq 0$ and $z \geq 0$, define the random event

$$A_I(\mu, z) = \{N_I \Delta_I^2 \leq \mu^2 z \text{ and } \Delta_I^* \leq 0.8 \min\{\mu, 1\}\}.$$

Theorem 4.3 *It holds for each* $\mu \geq 0$, $z \geq 0$ *and* $\beta \in (0,1)$

$$\mathsf{P}\{T_I \geq 2(1+2\mu)\ln(8N_I/\beta),\ A_I(\mu, z)\} \leq \beta.$$

As a consequence of this result, if the conditions $\Delta_I^* \leq 0.8\min\{\mu, 1\}$, $N_I\Delta_I^2 \leq \mu^2 z$ are fulfilled with probability one, then

$$\mathsf{P}\{T_I \geq 2(1+2\mu)\ln(8N_I/\beta)\} \leq \beta.$$

An extension of Theorem 4.2 to the nearly homogeneous case is also straightforward. Let a sequence $\{\beta_I\}$ satisfying $\sum_I \beta_I = \alpha$ be fixed. For every $I \in \mathcal{I}$ define $z_I = \ln(8N_I/\beta_I)$ and

$$A_I^*(\mu) = \{\Delta_{I'}^* \leq 0.8\min\{\mu, 1\},\ N_{I'}\Delta_{I'}^2 \leq \mu^2 z_{I'}\ \forall I' \in \mathcal{I}(I)\} = \bigcap_{I' \in \mathcal{I}(I)} A_I(\mu, z_{I'}).$$

Theorem 4.4 *For any* $\mu \geq 0$, *if* $\lambda_I \geq 2(1+2\mu)\ln(8N_I/\beta_I)$ *for all intervals* $I \in \mathcal{I}$, *then for every* $I \in \mathcal{I}$

$$\mathsf{P}\{I \textit{ is rejected},\ A_I^*(\mu)\} \leq \alpha_I.$$

4.5.4 Quality of the adaptive volatility estimate

The result of Theorem 4.4 suggests the following definition of a "good" or "ideal" interval $I\!I$. We say that $I\!I$ is good if for some fixed $\mu \geq 0$ the event $A_{I\!I}^*(\mu)$ meets with a high probability. Note that this definition does not uniquely describe the "ideal" intervals, that is, there can be many good intervals. It follows from Theorem 4.11 and Lemma 4.5 in the Appendix that the corresponding "oracle" estimate $\tilde{\theta}_{I\!I}$ delivers with a high probability the quality of estimation of order $N_{I\!I}^{-1/2}$.

The adaptive volatility estimate $\widehat{\theta}$ at the time point n is defined as $\tilde{\theta}_{\widehat{I}}$ where $\widehat{I}$ is the selected interval of homogeneity with the right end-point n. Here we show that the quality of the adaptive estimate is not worse (in order) than the quality of any "ideal" estimate corresponding to an "ideal" choice of the interval $\widehat{I}$. Let an interval $I\!I = [n - m^*, n[\in \mathcal{I}$ be fixed such that the event $A^*_{I\!I}(\mu)$ meets with a high probability. It is straightforward to see that the corresponding estimate $\tilde{\theta}_{I\!I}$ delivers the accuracy of order $N_{I\!I}^{-1/2}$. We now aim to show that the adaptive estimate $\widehat{\theta}$ provides at least the same (in order) accuracy of estimation. In the next result we assume for the ease of exposition that the procedure is run with the maximal set $\mathcal{I}$ of all possible intervals I of length $N_I \geq m_0$.

Theorem 4.5 *Let $I\!I$ be a "good" interval, that is, for some fixed μ, the event $A^*_{I\!I}(\mu)$ meets with a positive probability. If $\lambda_I \geq 2(1+2\mu)\ln(8N_I/\beta_I)$ for all intervals $I \in \mathcal{I}$, then the adaptive estimate $\widehat{\theta}$ fulfills*

$$\mathsf{P}\left\{\left|\ln(\widehat{\theta}/\theta_0)\right| \geq C\sqrt{\lambda_{I\!I}/N_{I\!I}},\ A^*_{I\!I}(\mu)\right\} \leq \alpha_{I\!I}$$

where the constant C depends on the parameter ρ of the procedure only.

4.5.5 Accuracy of estimation when θ_t is smooth

Suppose that the volatility function θ_t smoothly varies with t. Then the result of Theorem 4.5 can be used to state the usual nonparametric rate of estimation of the function θ. Since the volatility model $R = \theta^{1/2}\xi$ has irregularity at $\theta = 0$, it is more convenient to work with the log-transform of θ. We suppose that $\ln\theta_t$ is Lipschitz at n with the constant L, that is,

$$|\ln\theta_n - \ln\theta_t| \leq L\,|(n-t)/n|\,. \tag{4.7}$$

This condition implies for every interval $I = [n-m, n[$ that

$$\Delta^*_I \leq \max_{t\in I}|\theta_n/\theta_{n-m} - 1| \leq 2Lm/n$$

for $m \leq n/2$, that is, the conditions entering in the definition of the event $A^*_I(\mu)$ are satisfied almost sure with $\mu^2 = m(2Lm/n)^2/\ln(m/\beta_I)$. Selecting

$m = (n/2L)^{2/3}$ provides this condition with $\mu = 1$. Due to Theorem 4.5, the adaptive estimate $\widehat{\theta}$ ensures with a high probability the quality of estimation

$$\bigl|\ln(\widehat{\theta}) - \ln(\theta_0)\bigr| \le C\sqrt{\lambda_I/m} \le C\sqrt{\lambda_I}(2L/n)^{1/3}$$

which yields the usual nonparametric rate of estimation for smoothness degree one. We conclude with the following result:

Theorem 4.6 *Let θ_t fulfill (4.7) with probability one. Let also $\lambda_I \ge 6\ln(8N_I/\beta_I)$ for all intervals $I \in \mathcal{I}$. Define $m = [(n/2L)^{2/3}]$ and $I = [n-m, n[$. Then it holds for the adaptive estimate $\widehat{\theta}$*

$$\mathsf{P}\left\{\bigl|\ln(\widehat{\theta}) - \ln(\theta_0)\bigr| \ge C\sqrt{\lambda_I}(2L/n)^{1/3}\right\} \le \alpha_I$$

where the constant C depends on the parameter ρ of the procedure only.

4.5.6 Change point model

A *change point* model is described by a sequence $\nu_1 < \nu_2 < \ldots$ of Markov moments with respect to the filtration $\mathcal{F}_t$ and by values $\sigma_{(1)}, \sigma_{(2)}, \ldots$ where each $\sigma_{(k)}$ is $\mathcal{F}_{\nu_k}$-measurable. By definition $\sigma_t = \sigma_{(k)}$ for $\nu_k \le t < \nu_{k+1}$ and σ_t is constant for $t < \nu_1$. This is an important special case of the model (4.1). It is worth mentioning that any volatility process σ_t can be approximated by such a change point model. For this special case, the above procedure has a very natural interpretation: when estimating at the point n we search for a largest interval of the form $[n-m, n[$ which does not containing a change point. This is done via testing for a change point within the interval candidate $I = [n-m, n[$.

The construction of the procedure automatically provides the prescribed level of the first kind error probability (probability of a "false alarm"). In this section we aim to show that the procedure delivers a near optimal quality of change point detection. The quality (sensitivity) of a change point procedure is usually measured by the mean delay between the occurrence of the change points and its detection.

To study this property of the proposed method, we consider the case of estimation at a point n immediately after a change point ν. The 'ideal' choice $I\!I$ of the interval of homogeneity is clearly $I\!I = [\nu, n[$. Theorem 4.5 claims that the quality of the estimation of θ_n by our adaptive procedure is essentially the same as if we knew the latest change point ν a priori. In this section we present a more detailed analysis of the change point model. In particular, we show that the change point will be detected in an early stage of the procedure provided that the magnitude of the change is sufficiently large.

Denote $m^* = |I\!I|$, that is, $m^* = n - \nu$. Let I be a larger interval containing the change point ν, that is, $I = [\nu - m, n[= [n - m^* - m, n[$ for some m, so that $|I| = m + m^*$, and let θ (resp. θ') denote the value of parameter θ_t before (resp. after) change point ν. To simplify the exposition we suppose below that $m = m^*$. An extension to the case when m/m^* is bounded away from zero and infinity is straightforward. We now aim to show that such an interval I will be rejected with a high probability. It suffices to check that one particular test corresponding to $\tau = \nu$ rejects the hypothesis, that is, $T_{I,\nu} \geq \lambda_I$ with a high probability. The construction of the test statistic $T_{I,\nu}$ and Lemma 4.2 from the Appendix suggest the following measure of change from θ to θ':

$$d^2(\theta, \theta') = K(\theta, \theta_I) + K(\theta', \theta_I)$$

where $\theta_I = (\theta + \theta')/2$.

Theorem 4.7 *Let $\theta_t = \theta$ before the change point at ν and $\theta_t = \theta'$ after it. Let $I = [\tau - 2m^*, n[$ with $m^* = n - \nu$. There exists an absolute constant C_1 such that the condition*

$$d(\theta, \theta') \geq (1 + C_1)\sqrt{\lambda_I / m^*} \tag{4.8}$$

implies

$$\mathsf{P}\,(I \text{ is not rejected}) \leq \beta_I\,.$$

The result of Theorem 4.7 delivers some additional information about the sensitivity of the proposed procedure to change points. One possible question is about the minimal delay m^* between the change point ν and the first moment n when the procedure starts to indicate this change point by selecting an interval of type $I\!I = [\nu, n[$. Due to Theorem 4.7, the change will be "detected" with a high probability if (4.8) meets. With fixed $\theta \neq \theta'$, condition (4.8) is fulfilled if m^* is larger than a prescribed constant, that is, we need only a finite number of observations to detect a change point. In general, m^* should be of order $d^{-2}(\theta, \theta') \asymp |\theta - \theta'|^{-2}$, if the size of the change becomes small.

Finally we discuss the quality of estimating the location of the change point by the presented procedure. Without loss of generality we can consider the change point model with only one change at a point ν and suppose that for the interval candidate I the point ν belongs to the set of testing points $\mathcal{T}_I$. We know from the previous result that if the size of the change is sufficiently large, then the procedure detects (with a high probability) a change point in the sense that the test statistic $T_{I,\tau}$ with $\tau = \nu$ fulfills $\mathsf{P}(T_{I,\tau} \geq \lambda_I) \approx 1$. Now we are interested to evaluate how precise our procedure estimates the location of the change point. Recall that the estimated location $\widehat{\nu}$ maximizes $T_{I,\tau}$ over all $\tau \in \mathcal{T}_I$. Here we want to show that the estimated location of the change point differs from the true location ν in typical situation at most by a finite number m.

Theorem 4.8 *Let $\theta_t = \theta$ before the change point at ν and $\theta_t = \theta'$ after it. Let I be such that $\nu \in \mathcal{T}_I$. There exists an absolute constant C_2 such that if*

$$K\{\theta, (\theta + \theta')/2\} \geq C_2 \lambda_I / m \quad \textit{and} \quad K\{\theta', (\theta + \theta')/2\} \geq C_2 \lambda_I / m$$

then

$$\mathsf{P}(|\widehat{\nu} - \nu| > m) \leq \alpha_I .$$

It is worth mentioning that the conditions $K\{\theta, (\theta + \theta')/2\} \geq C_2 \lambda_I / m$ and $K\{\theta', (\theta + \theta')/2\} \geq C_2 \lambda_I / m$ are asymmetric w.r.t. θ, θ'. Namely, if

$\theta > \theta'$ then $K\{\theta, (\theta + \theta')/2\} < K\{\theta', (\theta + \theta')/2\}$. This implies that the change from low to high volatility is easier to detect than the change for high to low volatility. All these issues are in agreement with the theory of change point detection, see, e.g. Pollak (1985) and Brodskij and Darkhovskij (1993), and with our numerical results from Section 4.6.

4.5.7 Extension to the multiple volatility modeling

The most of results presented so far, can be extended to the case of multiple volatility modeling, cf. Härdle et al. (2003). However, this extension is not straightforward. All the previous results heavily use the deviation bound for the fitted likelihood $\widehat{L}_I$ given in Theorem 4.10 of the Appendix. The proof of this bound essentially utilizes that the parameter space is one-dimensional. This allows to establish nice non-asymptotic bounds for the considered test statistics under the homogeneity and near homogeneity.

Here we state only one result for the multiple volatility case, namely, the bound for the "false alarm" probability under homogeneity. All the remaining results can be extended in a similar way. We assume the model (4.3) and consider the locally homogeneous situation, when the volatility matrix Σ_t is constant within the consider interval I. The aim is to extend to the multiple case the statement of Theorem 4.1 which will be effectively used for tuning the parameter λ_I of the procedure.

Theorem 4.9 *Let $I \in \mathcal{I}$ and Σ_t is constant for all $t \in I$. Then it holds for every $z \geq 0$*

$$\mathsf{P}\left(T_I \geq 2z\right) \leq 2N_I^{C_d} e^{-z}$$

where the constant C_d depends on the dimension d only. In particular, for every $\beta \in (0,1)$, with $z = \ln(N_I^{C_d}/\beta)$, it holds

$$\mathsf{P}\left\{T_I \geq 2\ln(2N_I^{C_d}/\beta)\right\} \leq \beta.$$

4.6 Simulated results and applications

This section illustrates the performance of the proposed local change point detection (LCPD) procedure by means of some simulated data sets and applications to real data. We aim to show that the theoretical properties of the method derived in the previous section are confirmed by the numerical results. We focus especially on the two main features of the method: stability under homogeneity and sensitivity to changes of volatility. We also discuss in greater detail the problem of parameter tuning for our procedure.

4.6.1 Parameter tuning

Here we specify the procedure which is applied for both, simulated study and applications. The parameter setting is almost identical for the univariate and multivariate cases. In this section we discuss the univariate case, and the differences in the parameter choice for the multivariate case are given in Section 4.6.3.

The family of tested intervals I_k on which the procedure is performed is defined in the following way:

$$I_k = [n - m_k\,, n[\qquad \text{where} \qquad m_k = [m_0 c^k] \qquad \text{for} \qquad k = 0, 1, 2, \ldots . \quad (4.9)$$

Here $[x]$ means the integer part of x. The value m_0 characterizes the length of the smallest tested interval and together with the choice of ρ it determines the length of the smallest interval which can be accepted. Note that for fixed m_0 and ρ the procedure involves the estimation of the volatility over an interval J of length $m_0\rho$. Therefore this value should not be too small. We apply $m_0 = 15$ and $\rho = 1/3$ for our simulated examples, leading to $m_0\rho = 5$. The choice of a larger m_0 slightly decreases the sensitivity of the procedure to changes in volatility but it improves the stability and robustness of the estimator. For real (financial) data the choice $m_0 = 30$ or even $m_0 = 60$ can be recommended.

The parameter c controls the growth rate of tested intervals I_k as it can be seen by (4.9). The largest admissible value is $c = 2$, that is, every interval

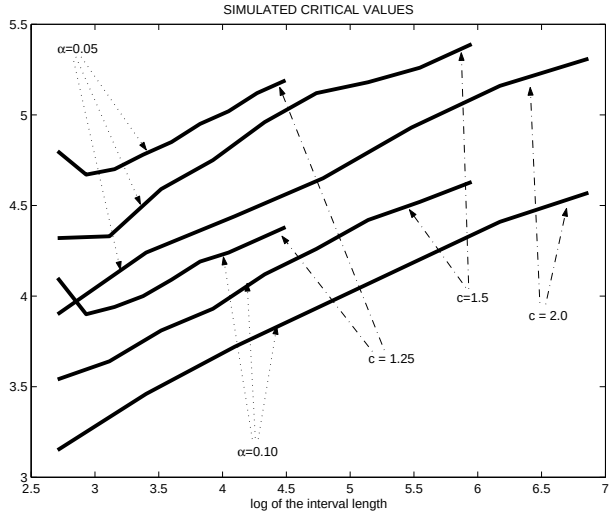

Figure 4.1: Critical values computed by simulation for different values of c and α.

I_k is twice as larger than I_{k-1}. Selecting a large c results in a rapid choice of the intervals I_k. This can lead to the situation that at some iteration an interval I_k with more than one change point can be considered, and our change point analysis may break down. Selecting a c close to one reduces the probability of such events but the payment for multiple testing in the critical values λ_I becomes larger which leads to a less sensitive procedure. However, as our simulation results and applications to real data show, the overall dependence of the estimation results on the parameter c is rather minor.

Finally we discuss the choice of the critical values λ_I. As suggested in Section 4.3.3, first we define the values β_{I_k} as

$$\beta_{I_k} = \alpha m_k^{-1} \left(\sum_{\ell=0}^{\infty} m_\ell^{-1} \right)^{-1} \approx \frac{\alpha(1-c^{-1})}{c^k},$$

and the corresponding value α_{I_k} is therefore $\alpha_{I_k} \approx \alpha(1-c^{-(k+1)})$. Then the critical values $\lambda^*_{I_k}$ are selected in order to provide the prescribed type-1-error

at level α_{I_k} for fixed values of c and α for every interval I_k, $k = 1, 2, \ldots$.

In our simulated examples we selected $c = 1.5$ and $\alpha = 0.05$. An extensive study which is not reported here showed that values of $c \in [1.1, 2]$ do not essentially affect the results of estimation neither on simulated, nor on real data, while different values of α lead to the usual trade off between type-1-error and type-2-error.

Figure 4.1 plots the corresponding critical values λ_I^* computed by simulation against the logarithm of the interval length $\ln(N_I)$. These values are in fact the empirical quantiles of the test statistics under the null hypothesis of homogeneity, i.e. constant volatility. Notice that under the null hypothesis the distribution of the test statistic does not depend on the particular value of the unknown volatility parameter. The computed values λ_I^* follow the linear relationship $\lambda_I^* = a + b \ln(N_{I_k})$. Moreover, the slope coefficient b is almost identical for all cases, and only the intercept depends on c and α.

The results of the approximated linear regressions are shown in Table 4.1. The slopes of all the regression lines are about 0.35 and only the intercepts vary across c's and α's. Therefore, to ease the implementation of the procedure we suggest to define the critical values in the following form:

$$\lambda_{I_k}(c, \alpha) = \lambda_0(c, \alpha) + 0.35 \ln(N_{I_k}), \tag{4.10}$$

where $\lambda_0(c, \alpha)$ corresponds to the intercept reported in Table 4.1.

4.6.2 Some simulated examples. Univariate case

Three different jump processes are simulated, whose relative jump magnitude is 3.00, 2.00 and 1.75 respectively. Each process is simulated and estimated 1000 times and the median and the quartiles of the estimates are plotted in Figure 4.2. We show the results for the final estimate $\widehat{\theta}$ and for the length of the selected interval $\widehat{I}$. One can see that if the size of the change is large enough, the procedure performs as if the location of the change were known.

Table 4.1: Intercept and slope for the linear regression of the log interval length on the critical values shown in Figure 4.1. For $c = 1.25$ we omit the first observation in the regression.

	$\alpha = 0.05$	$\alpha = 0.10$
$c = 2.00$	3.04 0.34	2.34 0.34
$c = 1.50$	3.34 0.35	2.58 0.35
$c = 1.25$	3.61 0.35	2.93 0.32

As one can expect, the sensitivity of the change point detection decreases when the magnitude of the jump becomes smaller. However, the accuracy of estimate of the volatility remains rather good even for small jumps that corresponds to our theoretical results.

The algorithm proposed is compared with the LAVE procedure from MS2004 with the optimized tuning parameters $\gamma = 0.5\,,\ M = 40\,,\ \lambda = 2.40\,.$ Figure 4.3 shows the quartiles of estimation for the two approaches for the model with the relative jump magnitude 3. One can see that the new procedure outperforms LAVE both with respect to the variance and to the bias of the estimator, especially for the points immediately after the changes.

Our simulation study has been done for the conditional normal model (4.1). We mentioned in Section 4.2.1 that this assumption is questionable as far as the real financial data is considered. To gain an impression about the robustness of the method against violation of the normality assumption we also simulated using i.i.d. innovations from the t_5-distribution with five degree of freedoms. The results are shown in Figure 4.4. As one can expect they are slightly worse than in the case of normal innovations, however, the procedure continues to work in a quite reasonable way. The sensitivity of the procedure remains as good as with normal innovations but a probability to reject a homogeneous interval became larger. This results in a higher variability of the estimated volatility.

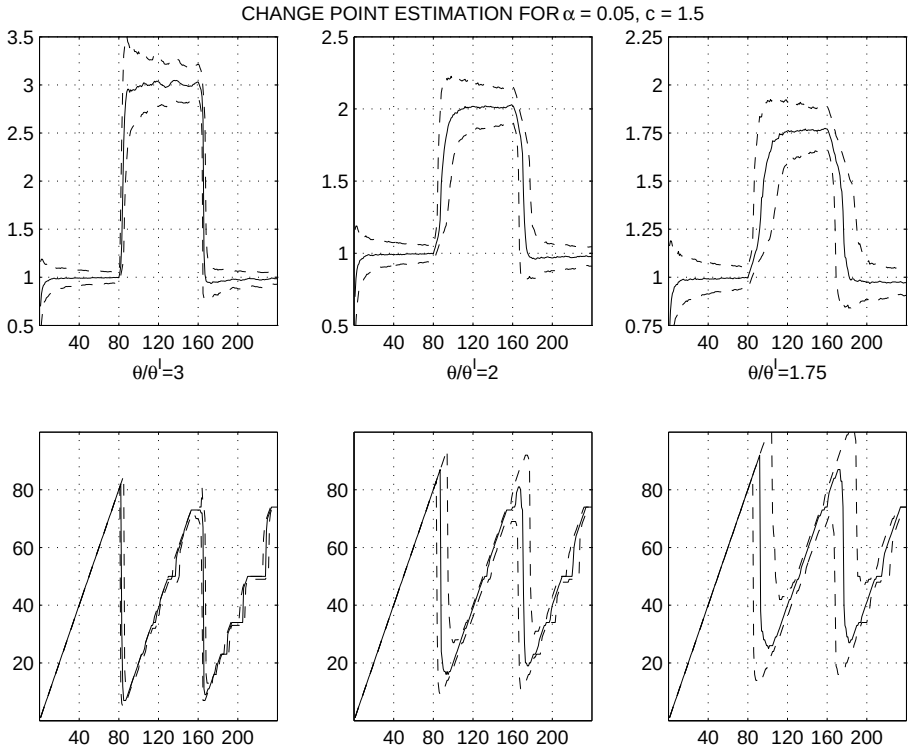

Figure 4.2: Pointwise median (solid line) and quartiles (dashed lines) for the estimates $\widehat{\theta}_t$ (top row) and the length of the selected interval $\widehat{I}_t$ for three jump processes with jumps of different magnitudes. The results are obtained with parameters $c = 1.5$ and $\alpha = 0.05$.

4.6.3 Some simulated examples. Multivariate case

The implementation of the estimator in the multivariate case is similar to the univariate case. In particular, for obtaining the critical values we again exploit the fact that the distribution of the likelihood ratio statistic under the hypothesis of homogeneity does not depend on the value of the covariance matrix, and therefore the critical values can be easily obtained by simulation.

In our implementation we again select $m_0 = 30$ and $c = 1.5$. Figure 4.5 shows the critical values as a function of the log interval length for a univari-

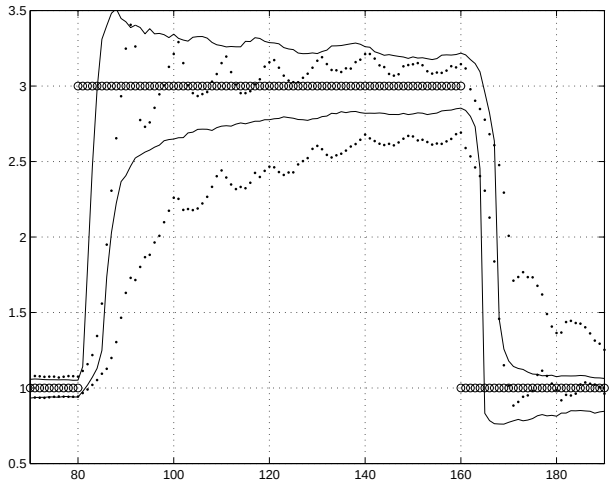

Figure 4.3: Comparison of the proposed estimator with the one from MS2004 for change-point model with $\theta/\theta' = 3$. Quartiles of $\widehat{\theta}$ for the new method (solid lines) and for the MS2004 (dotted lines).

ate, a bivariate and a trivariate system at a 5% level. Note that the curves indicate an approximately linear relationship and are almost parallel. This fact can be used to extrapolate critical values for larger intervals and for systems of a larger dimension.

The log length of the interval is regressed on the simulated critical values, and as suggested by Figure 4.5 we allow for three different intercepts but only one slope coefficient. The results of the regression are displayed in Table 4.2

Table 4.2: Linear approximation of the critical values

slope	intercept		
	univariate	bivariate	trivariate
0.31	3.39	6.33	9.94

and they suggest the use of the following linear rule for selecting critical values which keep the type-1-error approximately at a 5% and depend on the dimension of the system d and on the length of the currently tested

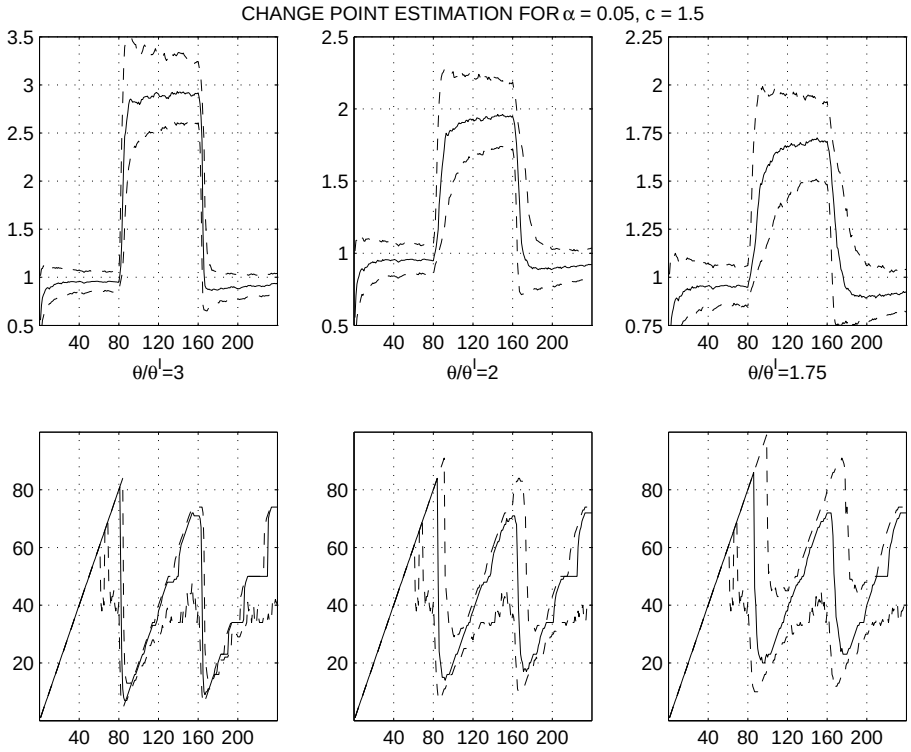

Figure 4.4: Estimation results with respect to jump processes with jumps of different magnitudes. The results are obtained with tuning parameters $c = 1.5$ and $\alpha = 0.05$. The conditional distribution is scaled student t_5 with 5 degrees of freedom.

interval $|I|$:

$$\lambda(d,\, |I|) = 3.3d + 0.31 \ln |I|. \tag{4.11}$$

Using the critical values given by (4.11) we apply the estimation procedure on simulated data.

We consider the following bivariate example. The correlation is set to

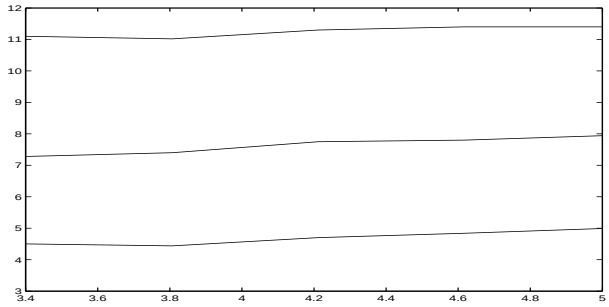

Figure 4.5: Critical values computed by simulation for $c = 1.5$ and $\alpha = 5\%$ for systems of different dimension. From the top: trivariate, bivariate and univariate.

zero and the volatilities are jump processes:

$$\begin{aligned}\sigma_{1t} &= 1 + 2\boldsymbol{I}_{\{t\in[101,200]\}} + \boldsymbol{I}_{\{t\in[301,400]\}} + \frac{3}{4}\boldsymbol{I}_{\{t\in[501,600]\}} \\ \sigma_{2t} &= 1 - \frac{2}{3}\boldsymbol{I}_{\{t\in[101,200]\}} - \frac{1}{2}\boldsymbol{I}_{\{t\in[301,400]\}} - \frac{3}{7}\boldsymbol{I}_{\{t\in[501,600]\}}.\end{aligned}$$

This system is simulated and estimated five hundred times and the average estimates of the volatilities, the median and the quartiles of the estimated interval of time homogeneity are plotted in Figure 4.6. As expected, the performance is very similar to the univariate case. The jump detection is quick and it is more accurate for larger jumps.

4.6.4 Volatility estimation for different exchange rate data sets

The volatility estimation is performed on a set of nine exchange rates, which are available from the web page of the US Federal Reserve. The data sets represent daily exchange rates of the US Dollar (USD) against the following currencies: Australian Dollar (AUD), British Pound (BPD) Canadian Dollar (CAD), Danish Krone (DKR), Japanese Yen (JPY), Norwegian Krone (NKR), New Zeeland Dollar (NZD), Swiss Franc (SFR) and Swedish Krone (SKR). The period under consideration goes from January 1st, 1990 to April 7th, 2000. For each time series we have 2583 observations. All selected time

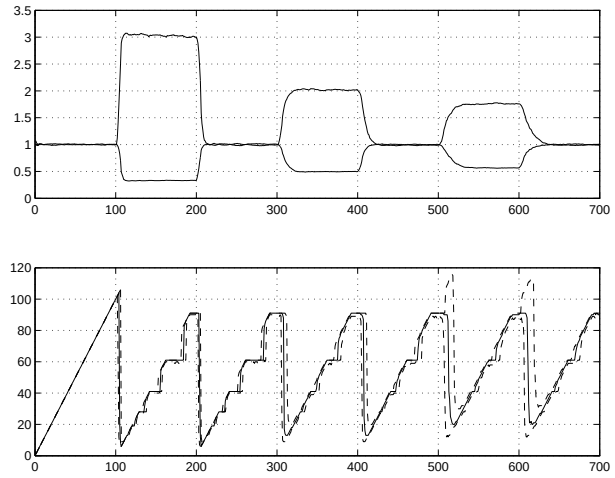

Figure 4.6: Simulation results for a diagonal bivariate process with jumps of different magnitude. Upper plot: pointwise median over 500 simulations for the two diagonal elements of the volatility matrix. Lower plot: median estimate of the interval of time homogeneity and quartiles.

series display excess kurtosis and volatility clustering.

Figure 4.7 and Figure 4.8 show the BPD/USD and JPY/USD exchange rate returns together with the volatility estimated with the parameters: $\alpha = 0.95\,,\ c = 1.5$ and $m_0 = 60\,.$ The choice of m_0 (which exceeds one used in the simulation) is made to improve the stability of the procedure against large shocks in the real data. The results of the estimation are in accordance with the data and the procedure seems to recognize changes in the underlying volatility process quickly.

The assumption of local homogeneity leads to the constant forecast $\widehat{\sigma}_t^2$ of the volatility σ_{t+h} for a small or moderate time horizon $h\,.$ This results in the following forecast of the conditional variance of the aggregated returns $R_{t+1}^2 + \ldots + R_{t+h}$:

$$V_{t,h}^{\mathtt{LCPD}} \stackrel{\mathrm{def}}{=} h\widehat{\sigma}_t^2,$$

with h being the forecast horizon.

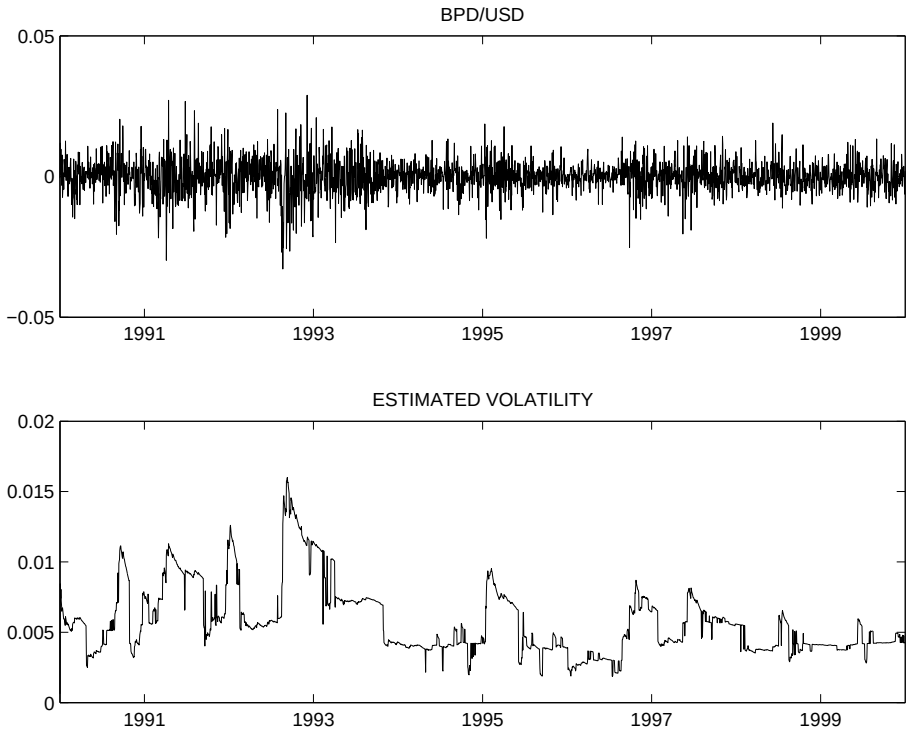

Figure 4.7: Returns and estimated volatility for the BPD/USD exchange rate.

In order to assess the performance of the proposed algorithm we compare its forecasting ability with the one of the GARCH(1,1) model, which represents one of the most popular parameterizations of the volatility process of financial time series. The GARCH(1,1) model is described by the following equations:

$$R_t = \sigma_t \xi_t, \quad \sigma_t^2 = \omega + \alpha R_{t-1}^2 + \beta \sigma_{t-1}^2,$$
$$\alpha > 0, \quad \beta > 0, \quad \alpha + \beta < 1, \quad \xi_t \sim \mathrm{N}(0,1)\ \forall t.$$

The h-step ahead variance forecast of the GARCH(1,1) is given by:

$$\sigma_{t+h|t}^{2,\texttt{GARCH}} \stackrel{\text{def}}{=} \mathsf{E}_t R_{t+h}^2 = \bar{\sigma}^2 + (\alpha+\beta)^h (\sigma_t^2 - \bar{\sigma}^2),$$

where $\bar{\sigma}$ represents the unconditional volatility and $\mathsf{E}_t \xi$ means $\mathsf{E}(\xi|\mathcal{F}_t)$, see

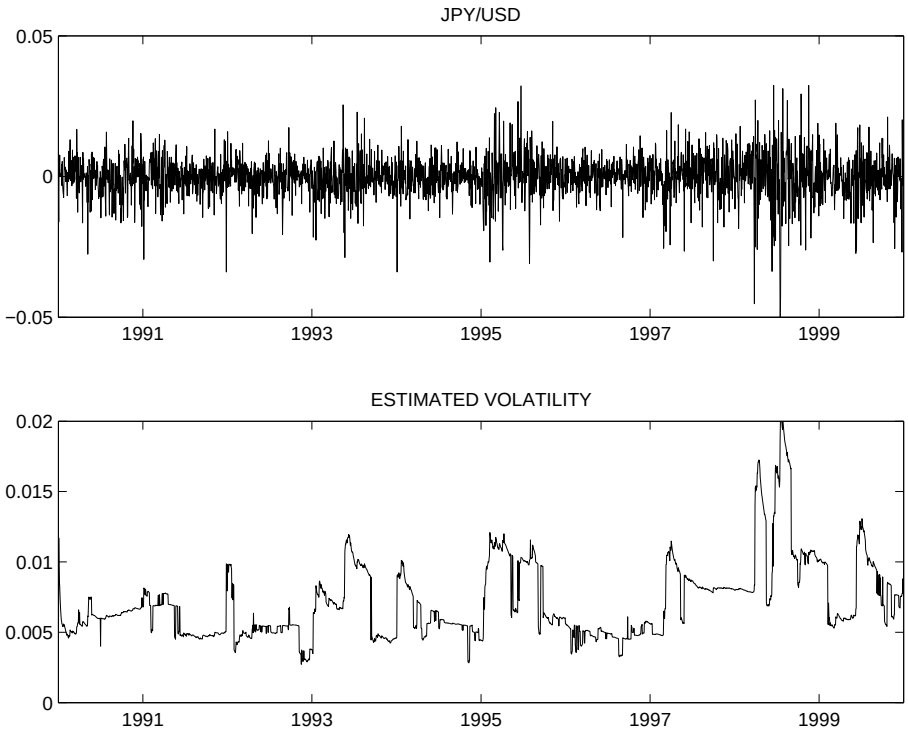

Figure 4.8: Returns and estimated volatility for the JPY/USD exchange rate.

Mikosch and Starica (2000a). Since the returns are conditionally uncorrelated, the conditional variance of the aggregated returns is given by the sum of the conditional variances:

$$V_{t,h}^{\texttt{GARCH}} \stackrel{\text{def}}{=} \mathsf{E}_t(R_{t+1} + \ldots + R_{t+h})^2 = \sum_{k=1}^{h} \mathsf{E}_t R_{t+k}^2 = \sum_{k=1}^{h} \sigma_{t+h|t}^{2,\texttt{GARCH}}.$$

The assumption of constant parameters for a GARCH(1,1) model over a time interval of the considered length of about 2500 time points can be too restrictive. We therefore considered a scrolling estimate, that is, for every date the preceding 1000 observations are used for estimation of the GARCH parameters and then the estimated parameters are used to forecast the variance at different horizons. This method is nonadaptive in the choice of the observation window but it takes advantage of a more flexible GARCH-modeling.

The suggested LCPD algorithm applies a very simple local constant modeling but benefits from a data-driven choice of the interval of homogeneity.

The quality of forecasting is measured by comparing the forecasts $V_{t,h}^{\mathtt{LCPD}}$ resp. $V_{t,h}^{\mathtt{GARCH}}$ with the realized volatility

$$\bar{V}_{t,h} \stackrel{\mathrm{def}}{=} R_{t+1}^2 + \ldots + R_{t+h}^2.$$

We apply the following mean square root error criterion (MSqE) for a time interval I:

$$\mathrm{MSqE}_I = \sum_{t \in I} |V_{t,h}^{\mathtt{LCPD}} - \bar{V}_{t,h}|^{1/2} \Big/ \sum_{t \in I} |V_{t,h}^{\mathtt{GARCH}} - \bar{V}_{t,h}|^{1/2}.$$

The MSqE is considered instead of the more common MSE for robustness reasons. Actually, in this way outliers are prevented from having a strong influence on the results. The MSqE is computed for six nonoverlapping intervals of 250 observations and the results are shown in Table 4.3. One can observe that both methods are comparable and that the relative performance depends on the particular situation at hand. For periods with stable volatility the LCPD forecast is clearly better, but for periods with high volatility variation the GARCH method is slightly preferable.

4.6.5 Analysis of standardized returns

Our model (4.1) assumes the standard normal innovations ξ_t. Many empirical researches argued that this assumption is too strong and often violated, see e.g. McNeil and Frey (2000). Here we briefly discuss this issue by looking at the standardized returns $\widehat{\xi}_t = R_t/\widehat{\sigma}_t$. The first observation is that even after standardization by the estimated variance, the density of standardized returns $\widehat{\xi}_t$ still displays tails which are fatter than the normal. We illustrate this effect in Figure 4.9 where the kernel estimate of the density of standardized returns $R_t/\widehat{\sigma}_t$ is plotted against the normal density and the scaled student t_5 density with 5 degrees of freedom. One can observe that the t-distribution delivers a much better approximation to the empirical density

Table 4.3: Relative forecasting performance MSqE on six consecutive time periods of 250 observations each.

	$h = 1$	**0.96**	**0.97**	**0.87**	1.09	1.15	1.02
AUD	$h = 5$	**0.98**	1.07	**0.83**	1.10	1.09	**0.99**
	$h = 10$	1.00	1.09	**0.85**	1.16	1.09	**0.98**
	$h = 1$	1.13	1.08	**0.74**	1.03	1.11	1.11
CAD	$h = 5$	1.13	1.11	**0.62**	1.04	1.08	1.15
	$h = 10$	1.15	1.12	**0.53**	1.02	1.06	1.16
	$h = 1$	**0.73**	**0.84**	**0.64**	1.06	**0.98**	1.02
BPD	$h = 5$	**0.66**	**0.85**	**0.60**	1.11	1.00	1.01
	$h = 10$	**0.61**	**0.84**	**0.55**	1.12	1.01	1.03
	$h = 1$	**0.87**	**0.99**	**0.76**	1.02	1.02	1.02
DKR	$h = 5$	**0.86**	1.03	**0.63**	1.04	1.04	1.01
	$h = 10$	**0.90**	1.01	**0.59**	1.05	**0.99**	1.03
	$h = 1$	1.10	1.16	**0.97**	1.16	1.15	1.03
JPY	$h = 5$	1.06	1.12	**0.85**	1.10	1.16	1.06
	$h = 10$	1.08	1.04	**0.80**	1.18	1.16	1.06
	$h = 1$	**0.87**	1.00	**0.88**	1.10	1.21	**0.94**
NKR	$h = 5$	**0.86**	1.01	**0.85**	1.10	1.29	**0.93**
	$h = 10$	**0.90**	1.01	**0.79**	1.14	1.32	**0.95**
	$h = 1$	**0.93**	**0.99**	**0.92**	1.11	1.24	1.13
NZD	$h = 5$	**0.92**	1.00	**0.88**	1.14	1.15	1.19
	$h = 10$	**0.82**	1.04	**0.88**	1.17	1.16	1.21
	$h = 1$	**0.93**	1.05	**0.84**	**0.95**	**0.98**	1.03
SFR	$h = 5$	**0.96**	1.08	**0.82**	**0.94**	**0.97**	1.11
	$h = 10$	**0.92**	1.03	**0.75**	**0.96**	**0.94**	1.19
	$h = 1$	**0.94**	**0.85**	**0.79**	1.02	1.10	**0.94**
SKR	$h = 5$	**0.94**	**0.85**	**0.69**	1.01	1.10	**0.94**
	$h = 10$	**0.95**	**0.89**	**0.62**	1.07	1.18	**0.93**

of returns.

The volatility clustering effect, though, disappears after standardization and autocorrelations of squared returns are not significant any more, see Figure 4.10 for the case of BPD/USD returns. The other exchange rate examples deliver similar results. A short conclusion of this empirical study is

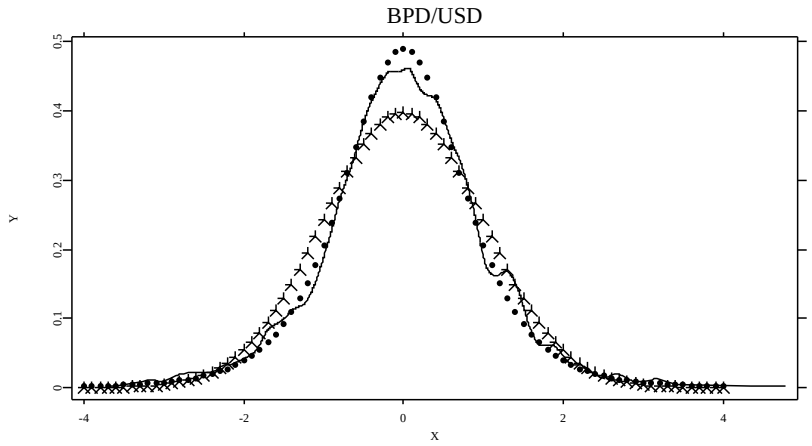

Figure 4.9: Kernel density estimate of exchange rate returns (solid line), normal density (x-line) and scaled student t_5 density with 5 degrees of freedom (dotted line) with fitted parameters for two exchange rate datasets.

that the standardized returns can be treated as i.i.d. random variables with a distribution whose tails are fatter than those of the normal distribution.

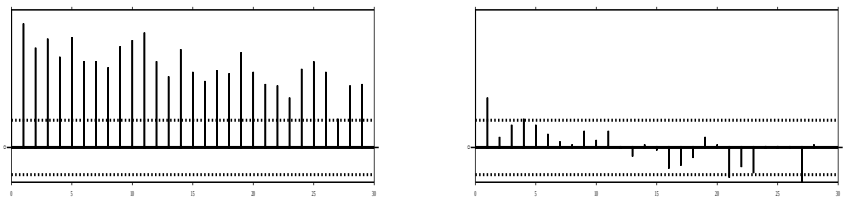

Figure 4.10: ACF of the absolute BPD/USD returns (upper plot) and of the standardized absolute BPD/USD returns (lower plot). Dotted straight line - the 95% significance level.

4.6.6 Multiple volatility estimation for exchange rate datasets

Now we apply the multiple LCPD procedure to the same set of nine exchange rates. Figure 4.11 shows the estimated diagonal elements of the covariance matrix. The upper plot shows the values estimated by the procedure described in Section 4.3. Similarly to the univariate case, these estimates have been constructed only from past observations and therefore be used for forecasting and other practical applications such as Value at Risk. Common movements, especially in the second part of the sample, can clearly be recognized, hinting that the volatilities of these processes are probably driven

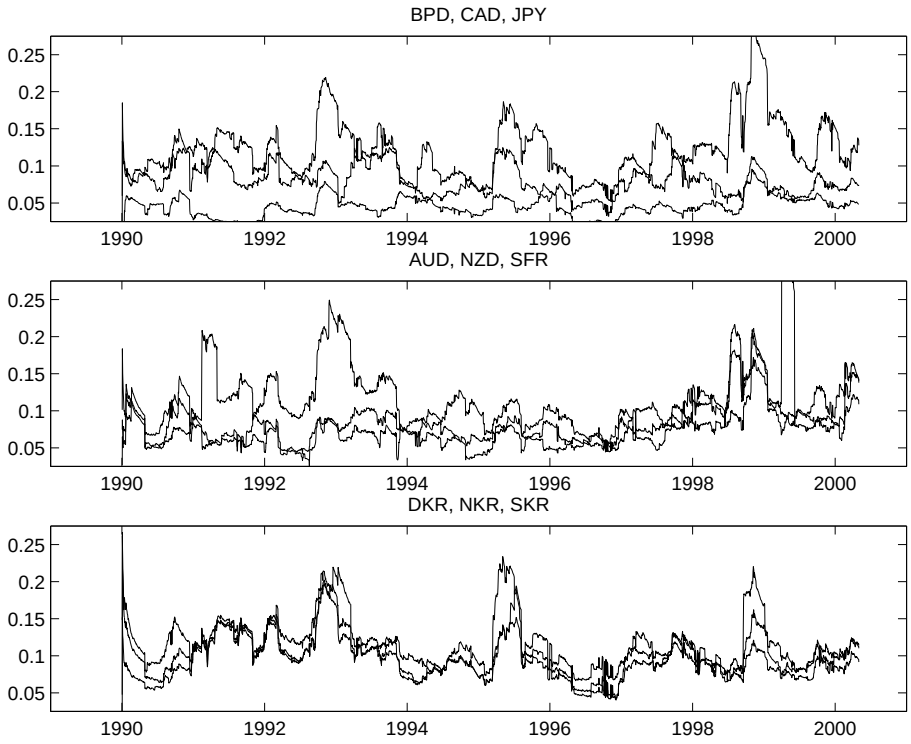

Figure 4.11: Adaptive estimate of the diagonal elements of the covariance matrix of nine exchange rate data sets. The scale is annualized volatility, i.e. we plot $\sqrt{250\widehat{\Sigma}_{ii}}$.

by some low dimensional common factor.

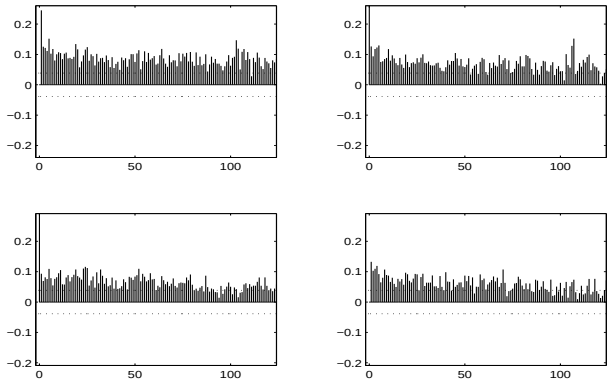

Figure 4.12: ACF for the absolute returns of NZD and AUD.

Figure 4.12 shows the multivariate autocorrelation function of the absolute returns for two exchange rate time series, New-Zealand Dollar and Australian Dollar, while, Figure 4.13 presents the multivariate autocorrelation function of the absolute standardized returns. Again, the standardized returns do not indicate significant correlations and autocorrelations. Note, that the autocorrelation of the absolute residuals has been almost completely smoothed away after standardizing by the estimated volatility matrix. The same conclusion holds for the other considered time series.

4.6.7 Application to Value at Risk. Univariate case

The Value at Risk (VaR) measures the extreme loss of a portfolio over a predetermined holding period with a prescribed confidence level $1-\alpha$. This problem can be reduced to computing the quantiles of the distribution of aggregated returns, see e.g. Fan and Gu (2003) for a recent overview of this topic.

Our modeling approach can easily be adapted to the VaR problem. Namely, one may forecast the 1% and 5% quantile of the next return R_{t+1} and of

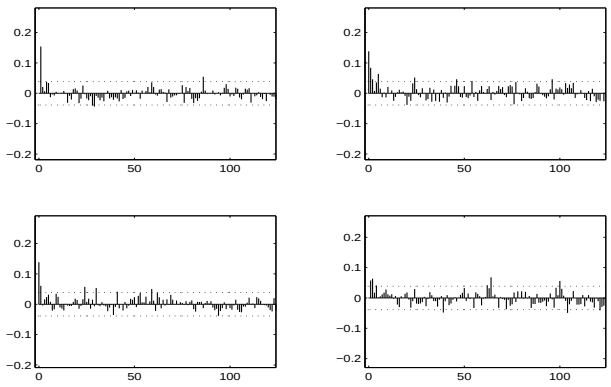

Figure 4.13: ACF for the absolute standardized returns of NZD and AUD.

the aggregated returns $R_{t+1} + \ldots + R_{t+h} = \ln(S_{t+h}/S_t)$ for each date t in the following way. The volatility parameter $\widehat{\sigma}_t$ is estimated from the historical data R_s for $s \leq t$ and one can consider different distributions for the innovations ξ_t. In our study we compare the Gaussian, the scaled student t_5-distribution with 5 degrees of freedom and the empirical distribution $\widehat{F}_t$ of the past empirical innovations $\widehat{\xi}_s$ for $s \leq t$, that is:

$$R_{t+h} = \widehat{\sigma}_t \xi_{t+h} \qquad \text{with} \qquad \xi_{t+h} \sim \mathrm{N}(0,1), \text{ or } \sqrt{5/3}\xi_{t+h} \sim t_5, \text{ or } \xi_{t+h} \sim \widehat{F}_t\,.$$

Similar approaches have been applied in McNeil and Frey (2000) with the use of the GARCH(1,1) model for estimating the volatility and extreme value theory for evaluating the distribution of returns, while Eberlein and Prause (2002) assume the Generalized Hyperbolic Distribution for the innovations.

In order to better interpret the results, we notice that the scaled t_5 distribution has higher 5% quantiles than the ones of the Gaussian at any of the considered horizons and lower 1% quantiles. Therefore the Gaussian distribution of innovations is more conservative for 5% quantiles while the opposite is true for 1% quantiles.

We apply the procedure to the set of nine exchange rates considered in Section 4.6.6 with about 2500 observations in each one. The frequency of overshooting the predicted quantile for the given realizations of the returns

is given in Table 4.4. The first 500 observations in every time series are taken as presample for estimating the parameters. Notice, that for the five and ten day horizon overlapping intervals of data are used as in Fan and Gu (2003).

According to the requirement of the regulators (BIS, 1996), a bank has to determine its capital requirements in order to cover from market risk proportionally to the 1% quantile of the distribution of the portfolio losses over a ten day horizon. Internal models calculating this quantile are regularly monitored. The coefficient of proportionality is set to 3 for models whose performance is satisfactory (green zone) and it can be increased up to 4 by a discretionary judgment of the regulators for models which appear to estimate the quantile imprecisely (yellow zone). While, if the model performance is considered very poor, the coefficient is automatically increased to 4 (red zone).

The official criterion for the evaluation of an internal model is the statistical significance of the 1% quantile estimates of the portfolio loss distribution over a one day horizon. The prescribed procedure, called backtesting, checks, whether the observed frequency of days out of the last 250, for which the losses were larger than the value computed by the prescribed VaR procedure does not significantly deviates from the nominal level 0.01, see Deutsch (2001). Every procedure is classified as green, yellow and red. The green zone means that the empirical frequency is in agreement with the nominal probability 0.01. The yellow zone begins at the point such that the probability of exceptions for the tested VaR procedure exceeds the value 0.01 with 95% confidence interval. One can easily verify that such probability corresponds to 5 or more exceptions out of 250 days, that is, the frequency of exceptions equals 2%. Similarly, the red zone corresponds to the 99.99% level evidence that the tested procedure does not provide the required probability of exceptions. For a sample of 250 observations, this corresponds to 10 exceptions, or equivalently, 4% frequency of overshooting the VaR value.

The comparison of these requirements with our results presented in Table 4.4 shows that on average none of the procedures we tried is in the red

zone, and that the procedure using e.d.f. for the residuals is always in the green zone. The use of the student t_5 distribution also allows to get the green zone results for most of the examples, while the procedure with Gaussian innovations is often in the yellow zone.

We conclude that the use of the t_5 distribution for the innovations slightly improve the results and the VaR quality is acceptable for both Gaussian and scaled student quantiles, while the application of the empirical distribution of the residuals leads to almost perfect fit of the prescribed quantiles for all considered time horizons.

Table 4.4: Percentage of overshooting the prescribed VaR level for nine series of exchange rates for nominal quantile levels 1% and 5%, three different distributions of innovations and time horizon $h = 1, 5, 10$.

	1% quantile									5% quantile								
	Gaussian			student t_5			e.d.f.			Gaussian			student t_5			e.d.f.		
h	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10
AUD	2.3	2.7	2.2	1.9	2.4	2.1	0.7	0.2	0.7	5.5	5.9	6.3	6.3	6.0	6.4	3.9	3.1	2.8
CAD	1.7	1.6	0.9	1.0	1.2	0.8	1.5	1.7	1.9	4.7	5.3	4.5	5.2	5.6	4.7	5.6	7.2	6.6
BPD	2.4	2.5	2.4	1.6	2.3	2.1	1.0	0.9	0.9	5.3	7.1	6.5	6.0	7.4	6.7	4.2	4.2	3.2
DKR	2.4	2.2	1.9	1.7	1.6	1.6	0.9	1.5	1.6	5.8	6.5	6.5	6.5	6.5	6.6	4.7	5.3	5.7
JPY	2.7	3.3	3.5	1.9	3.1	3.2	1.0	1.4	1.3	5.5	7.5	8.5	6.0	7.7	8.5	4.4	4.9	4.8
NKR	2.0	1.9	1.4	1.3	1.5	1.2	0.8	1.4	1.3	5.5	5.7	6.0	6.3	6.0	6.1	4.8	4.4	5.0
NZD	2.8	2.7	3.1	2.1	2.5	2.7	0.7	0.7	1.0	5.1	6.1	6.7	5.5	6.2	6.9	4.0	4.7	4.1
SFR	1.8	2.0	2.5	1.2	1.4	2.3	1.0	1.3	1.6	5.8	6.0	6.0	6.4	6.0	6.1	4.5	5.0	5.8
SKR	1.7	1.3	1.0	1.2	1.1	0.8	0.7	1.2	1.4	6.2	5.7	4.9	6.7	6.2	5.1	4.3	4.4	4.8

4.6.8 Value at Risk for multiple time series

Here we illustrate the same approach for the portfolio containing several assets. Suppose the portfolio consists of N assets with the vector of allocations $\psi \in I\!R^N$. The allocations are usually time dependent, that is, at time t we

have $\psi_i = \psi_i(t)$ units of the assets i in the portfolio, $i = 1, \ldots, N$.

The goal is to estimate the Value at Risk for the whole portfolio. The estimated covariance matrix $\widehat{\Sigma}$ is one of the most important input for this problem. We present two possible approaches for computing VaR. One is based on the assumption that the joint distribution of the returns is multivariate normal, while the other estimates the quantiles from the empirical distribution of the returns standardized by the estimated covariance matrix and therefore can be interpreted as a version of the historical simulation approach.

Formally, for a given fixed allocation ψ, the VaR for an h day horizon and a probability level α of the portfolio $\psi^\top S_t$, is defined as the α-quantile of the distribution of the changes in the portfolio value:

$$\mathsf{P}\{\psi^\top (S_{t+h} - S_t) < \mathrm{VaR} | S_t\} = \alpha.$$

For a given S_t, the estimation of the VaR of the portfolio changes can be obtained if one is able to determine the conditional distribution of the sum of the returns:

$$\sum_{i=1}^{N} \psi_i \left(\frac{S_{t+h,i} - S_{t,i}}{S_{t,i}} \right).$$

For small h the above expression can be conveniently approximated by the sum of the log-returns:

$$\sum_{i=1}^{N} \psi_i \{\ln(S_{t+h,i}) - \ln(S_{t,i})\} = \sum_{i=1}^{N} \psi_i (R_{t+1,i} + \ldots + R_{t+h,i}) = \sum_{u=1}^{h} \psi^\top R_{t+u}\,,$$

so that for the computation of the VaR we can exploit the properties of the log-returns. In particular, if we assume that the returns are normally distributed with the covariance matrix Σ_t, then

$$\psi^\top (R_{t+1} + \ldots + R_{t+h}) \sim \mathrm{N}(0, h\psi^\top \Sigma_t \psi).$$

This suggests to compute the VaR using the quantiles of the $\mathrm{N}(0, h\psi^\top \widehat{\Sigma}_t \psi)$.

Similar to the univariate case, the normal distribution assumption is not very accurate for modeling the tails of the distribution of financial returns. One possibility to cope with this problem consists in fitting to the returns a distribution, which can better approximate the tail behavior of the data as in Eberlein and Prause (2002). We consider however another strategy, which consists in estimating the quantiles from the empirical distribution function of the standardized returns. Indeed, as shown in the previous Section, the standardized returns $\widehat{\xi}_t \stackrel{\text{def}}{=} \widehat{\Sigma}_{t-1}^{-1/2} R_t$ are approximately independent, therefore the quantile of the distribution of $\psi^\top \sum_{u=1}^{h} R_{t+u}$ can be estimated by the quantile of the empirical distribution function of $\psi^\top \widehat{\Sigma}_t^{\frac{1}{2}} (\widehat{\xi}_{t+1} + \ldots + \widehat{\xi}_{t+h})$.

The estimation of VaR is performed on a data set of nine exchange rates with 2583 daily observations. The first 82 observation are used to perform the first estimation of the covariance matrix and the following 500 observations are used to perform the first estimation of the empirical distribution function. On the last 2000 observations VaR is computed every day using all the past standardized residuals for the estimation of the empirical distribution function. For the sake of comparison, VaR is also computed under Gaussian assumption on the last 2000 observations. Table 4.5 reports the frequency with which the realization of the portfolio return is lower than the estimated quantile. We consider the values of $\alpha = 0.05$ and 0.01 and the horizons of 1, 5 and 10 trading days for two different portfolios. One portfolio has equal weights for each asset for the whole period, while the other portfolio has randomly generated weights from an uniform distribution on $[0,1]^9$ for each day.

Table 4.5: Value at risk for two portfolios with 9 assets

	1% quantile						5% quantile					
	Gaussian			e.d.f.			Gaussian			e.d.f.		
horizon	1	5	10	1	5	10	1	5	10	1	5	10
equally weighted portfolio	2.3	2.0	1.9	1.0	0.8	0.5	6.0	5.8	5.5	4.0	3.5	3.3
random uniform portfolio	2.1	2.2	2.1	0.9	0.7	0.7	6.3	6.4	5.7	4.1	3.5	3.2

As expected the Gaussian model slightly overestimates the quantiles (yellow zone) and the method which relies on the estimation of the empirical distribution function performs remarkably well in particular for the 1% quantile (green zone). For the 5% quantile the method seems to be slightly conservative.

4.7 Appendix

In this section we state some results about the properties of the log-likelihood in the time-inhomogeneous volatility model and present the proofs of the results stated in previous sections.

4.7.1 Some properties of the log-likelihood in the homogeneous case

Let $I = [n-m, n[$ be an interval from $\mathcal{I}$. Here we present some useful results about the properties of the log-likelihood $L_I(\theta)$ and the fitted log-likelihood $\widehat{L}_I = \max_\theta L_I(\theta) = L_I(\tilde{\theta}_I)$.

First we consider the homoskedastic situation when the volatility parameter is indeed constant within I. Denote $L_I(\theta, \theta') = L_I(\theta) - L(\theta')$ for any θ, θ'.

Theorem 4.10 *Let $\theta_t = \theta_0$ for all $t \in I$ where θ_0 is a constant or a random variable measurable w.r.t. $\mathcal{F}_{n-m-1}$. Then it holds for any θ and any $\lambda \geq 0$*

$$\mathsf{P}\left\{L_I(\theta, \theta_0) \geq \lambda\right\} \leq e^{-\lambda}$$

and

$$\mathsf{P}\left\{L_I(\tilde{\theta}_I, \theta_0) \geq \lambda\right\} \leq 2e^{-\lambda}.$$

This result will be proved as a part of a more general result which applies if the volatility process is nearly homogeneous within I.

4.7.2 Properties of the log-likelihood in the nearly homogeneous case

Suppose that there exists a value θ_0 measurable w.r.t. $\mathcal{F}_{n-m-1}$ such that the values Δ_I^* and Δ_I defined by the equations (4.6) are small with a high probability. Recall the notation $A_I(\mu, z) = \{N_I \Delta_I^2 \le \mu^2 z,\ \Delta_I^* \le 0.8 \min\{\mu, 1\}\}$.

Theorem 4.11 *Let $\mu \ge 0$. Then it holds for any θ and any $z \ge 0$*

$$\mathsf{P}\{L_I(\theta, \theta_0) \ge z + 2z\mu,\ A_I(\mu, z)\} \le 2N_I e^{-z}$$

and

$$\mathsf{P}\left\{L_I(\tilde{\theta}_I, \theta_0) \ge z + 2z\mu,\ A_I(\mu, z)\right\} \le 4N_I e^{-z}.$$

Proof Using standard technique one can easily reduce the results of the theorem to the case when the event $A(\mu, z)$ meets almost sure, so everywhere in the proof we assume that $\Delta_I^* \le 0.8 \min\{\mu, 1\}$ and $N_I \Delta_I^2 \le \mu^2 z$ with probability one.

The log-likelihood ratio can be represented as

$$L_I(\theta, \theta_0) = L_I(\theta) - L_I(\theta_0) = \left(\frac{1}{2\theta_0} - \frac{1}{2\theta}\right) S_I - \frac{N_I}{2} \ln(\theta/\theta_0).$$

with $S_I = \sum_{t \in I} R_t^2$.

Lemma 4.1 *For given z, there exist two values $\theta^* > \theta_0$ and $\theta_* < \theta_0$ depending on z, θ_0, N_I only such that*

$$\{L_I(\tilde{\theta}_I, \theta_0) \ge z\} \subseteq \{L_I(\theta^*, \theta_0) \ge z\} \cup \{L_I(\theta_*, \theta_0) \ge z\}.$$

Proof It holds

$$\begin{aligned}
\{L_I(\tilde{\theta}_I, \theta_0) \ge z\} &= \left\{\sup_{\theta} \{S_I(1/\theta_0 - 1/\theta) - N_I \ln(\theta/\theta_0)\} \ge 2z\right\} \\
&\subseteq \left\{S_I \ge \inf_{\theta > \theta_0} \frac{2z + N_I \ln(\theta/\theta_0)}{1/\theta_0 - 1/\theta}\right\} \cup \left\{-S_I \ge \inf_{\theta < \theta_0} \frac{2z + N_I \ln(\theta/\theta_0)}{\theta^{-1} - \theta_0^{-1}}\right\}.
\end{aligned}$$

It is straightforward to see that the function $f(u) = \{2z + N_I \ln(\theta/\theta_0)\} \,/\, \bigl(\theta_0^{-1} - \theta^{-1}\bigr)$ attains its minimum at some point $\theta^* > \theta_0$. Therefore

$$\left\{ S \geq \inf_{\theta > \theta_0} \frac{2z + N_I \ln(\theta/\theta_0)}{1/\theta_0 - 1/\theta} \right\} = \left\{ S \geq \frac{2z + N_I \ln(\theta^*/\theta_0)}{1/\theta_0 - 1/\theta^*} \right\} \subseteq \{ L_I(\theta^*, \theta_0) \geq z \}.$$

Similarly

$$\left\{ -S \geq \inf_{\theta < \theta_0} \frac{2z + N_I \ln(\theta/\theta_0)}{1/\theta - 1/\theta_0} \right\} \subseteq \{ L_I(\theta_*, \theta_0) \geq z \}$$

for some $\theta_* < \theta_0$.

Lemma 4.2 *For any* $z \geq 0$

$$\mathsf{P}\left(\breve{L}_I(\theta) \geq z \right) \leq e^{-z}.$$

where $\breve{L}_I(\theta) = \ln d\mathsf{P}_\theta / d\mathsf{P} = \sum_{t \in I} \bigl\{ \ell(R_t, \theta) - \ell(R_t, \theta_t) \bigr\}$. *Moreover, assuming the condition* $\Delta_I^* \leq 0.8 \min\{\mu, 1\}$ *and* $N_I \Delta_I^2 \leq \mu^2 z$, *it holds*

$$\mathsf{P}\left(-\breve{L}_I(\theta) \geq 2\mu z \right) \leq e^{-z}$$

and

$$\mathsf{P}\left\{ \pm \left(\breve{L}_I(\theta) + \mathcal{K}_I(\theta) \right) \geq 2\mu z \right\} \leq e^{-z}$$

where $\mathcal{K}_I(\theta) = \sum_{t \in I} K(\theta_t, \theta)$.

Proof Since $\breve{L}_I(\theta)$ is the log-likelihood, it obviously holds

$$\mathsf{P}\left\{ \breve{L}_I(\theta) \geq z \right\} \leq e^{-z} \mathsf{E} \exp \breve{L}_I(\theta) = e^{-z}$$

and the first assertion follows. Next,

$$\ln \mathsf{P}\left\{ -\breve{L}_I(\theta) \geq 2\mu z \right\} \leq -2z + \ln \mathsf{E} \exp\left\{ -\mu^{-1} \breve{L}_I(\theta) \right\}.$$

Since R_t is progressively measurable and θ_t is predictable w.r.t. $\mathcal{F}_t$, it holds by Lemma 4.4 below

$$\begin{aligned} \mathsf{E} \exp\left\{ -\mu^{-1} \breve{L}_I(\theta) \right\} &= \mathsf{E} \prod_{t \in I} \mathsf{E} \exp\left[\mu^{-1} \bigl\{ \ell(R_t, \theta_t) - \ell(R_t, \theta) \bigr\} | \mathcal{F}_{t-1} \right] \\ &\leq \mathsf{E} \prod_{t \in I} \exp\bigl(|\theta_t/\theta_0 - 1|^2 / \mu^2 \bigr) \leq \mathsf{E} \exp\bigl(N_I \Delta_I^2 / \mu^2 \bigr) \leq e^z \end{aligned}$$

and the second assertion follows. Similarly

$$\ln \mathsf{P}\Big\{\breve{L}_I(\theta) + \mathcal{K}_I(\theta) \geq 2\mu z\Big\} \leq -2z + \ln \mathsf{E} \exp\Big[\mu^{-1}\big\{\breve{L}_I(\theta) + \mathcal{K}_I(\theta)\big\}\Big]$$

and

$$\begin{aligned} \mathsf{E} e^{\left(\breve{L}_I(\theta)+\mathcal{K}_I(\theta)\right)/\mu} &= \mathsf{E} \prod_{t\in I} \mathsf{E} \exp\big[\mu^{-1}\big\{\ell(R_t,\theta) - \ell(R_t,\theta_t) + K(\theta_t,\theta)\big\}|\mathcal{F}_{t-1}\big] \\ &\leq \mathsf{E} \prod_{t\in I} \exp\big(|\theta_t/\theta_0 - 1|^2/\mu^2\big) \leq \mathsf{E} \exp\big(N_I \Delta_I^2/\mu^2\big) \leq e^z. \end{aligned}$$

A bound for $-\breve{L}_I(\theta) - \mathcal{K}_I(\theta)$ can be proved similarly.

Now we are prepared to complete the proof of the theorem. Indeed, $L_I(\theta,\theta_0) = \breve{L}_I(\theta) - \breve{L}_I(\theta_0)$ and Lemma 4.2 implies

$$\mathsf{P}\left\{L_I(\theta,\theta_0) \geq \lambda + 2\mu\lambda\right\} \leq \mathsf{P}\Big\{\breve{L}_I(\theta) \geq \lambda\Big\} + \mathsf{P}\Big\{-\breve{L}_I(\theta_0) \geq 2\mu\lambda\Big\} \leq 2e^{-\lambda}$$

and the result of the theorem follows by Lemma 4.1.

4.7.3 Proof of Theorem 4.3

Similarly to the proof of Theorem 4.11 we reduce the general situation to the case when the conditions $N_I\Delta_I^2 \leq \mu^2 \ln(N_I/\beta)$ and $\Delta_I^* \leq 0.8\mu$ are fulfilled almost surely. This automatically yields $N_{I'}\Delta_{I'}^2 \leq \mu^2 \ln(N_I/\beta)$ and $\Delta_{I'}^* \leq 0.8\mu$ for all the subintervals J of I.

Let some point $\tau \in \mathcal{T}_I$ be fixed with the corresponding subintervals J and J^c. Then

$$T_{I,\tau} = L_J(\tilde{\theta}_J,\theta_0) + L_{J^c}(\tilde{\theta}_{J^c},\theta_0) - L_I(\tilde{\theta}_I,\theta_0) \leq L_J(\tilde{\theta}_J,\theta_0) + L_{J^c}(\tilde{\theta}_{J^c},\theta_0).$$

Here $L_J(\theta,\theta_0)$ means $L_J(\theta) - L_J(\theta_0)$. We also used that $L_I(\tilde{\theta}_I,\theta_0) \geq L_I(\theta_0,\theta_0) = 0$.

Now, it holds by Theorem 4.11 that

$$\mathsf{P}\Big\{L_J(\tilde{\theta}_J,\theta_0) \geq (1+2\mu)\ln(8N_I/\beta)\Big\} \leq 4\exp\big\{-\ln(8N_I/\beta)\big\} \leq \beta/(2N_I).$$

Similarly one can bound $L_{J^c}(\tilde{\theta}_{J^c},\theta_0)$, so that

$$\mathsf{P}\left\{T_{I,\tau} \geq 2(1+2\mu)\ln(8N_I/\beta)\right\} \leq \beta/N_I.$$

This implies the result of the theorem because the number of testing intervals J does not exceed N_I.

4.7.4 Proof of Theorem 4.5

The next statement is the key step of the proof. Let an interval $I = [n-m, n[$ be accepted by the procedure. We aim to show that there exists $\tau \in \mathcal{T}_I$ such that the adaptive estimate $\widehat{\theta}$ fulfills with $J = [n-\tau, n[$

$$\left|\ln(\widehat{\theta}/\tilde{\theta}_J)\right| \leq C_0\sqrt{\lambda_I/N_I} \tag{4.12}$$

for some constant C_0 depending on the parameter ρ of the procedure. Indeed, the definition of the procedure implies for every accepted interval I and every point $\tau \in [n-m+m\rho, n-m\rho]$ that

$$T_{I,\tau} = N_J K(\tilde{\theta}_J, \tilde{\theta}_I) + N_{J^c} K(\tilde{\theta}_{J^c}, \tilde{\theta}_I) \leq \lambda_I\,.$$

Since the Kullback-Leibler information K is nonnegative, this also implies $K(\tilde{\theta}_J, \tilde{\theta}_I) \leq \lambda_I/N_J$. Let now $\widehat{I}$ be the selected interval of the form $[n-\widehat{m}, n[$. Define $n_0 = \widehat{m}$, $n_j = [n_{j-1}/2]$, $j = 1, 2, \ldots$. Because $\rho \leq 1/3$, there is some $j^* \geq 0$ such that $n_{j^*} \in [m\rho, m(1-\rho)]$. Now consider the sequence of intervals $U_j = [n-n_j, n[$ for $j = 0, \ldots, j^*$. Since, for every $j \geq 1$, the interval U_{j-1} is accepted and U_j is one of its testing intervals, it holds $K(\tilde{\theta}_{U_j}, \tilde{\theta}_{U_{j-1}}) \leq \lambda_{U_{j-1}}/N_{U_j} \leq 2\lambda_{U_{j-1}}/N_{U_{j-1}}$ and, by Lemma 4.5 below, it holds $\left|\ln(\tilde{\theta}_{U_j}/\tilde{\theta}_{U_{j-1}})\right| \leq \sqrt{12\lambda_{U_{j-1}}/N_{U_{j-1}}}$. This yields for $\widehat{\theta} = \tilde{\theta}_{U_0}$

$$\left|\ln(\widehat{\theta}/\tilde{\theta}_{I(j^*)})\right| \leq \sum_{j=1}^{j^*} \sqrt{12\lambda_{U_{j-1}}/N_{U_{j-1}}} \leq 8\sqrt{\lambda_{U_{j^*}}/N_{U_{j^*}}}.$$

Here we have used that $N_{U_{j-1}} \geq 2N_{U_j}$ for all $j \leq j^*$ and that λ_I grows at most logarithmically with N_I. It remains to note that $N_{U_{j^*}} \geq \rho N_I$ and (4.12) follows.

By Theorem 4.4 the interval $I\!I$ will be accepted with a high probability. Moreover, in the proof of Theorem 4.4 we showed that for all testing intervals J holds with a high probability $L_J(\tilde{\theta}_J, \theta_0) = N_J K(\tilde{\theta}_J, \theta_0) \leq \lambda_{I\!I}/2$, which implies by Lemma 4.5 that

$$\left|\ln(\tilde{\theta}_J/\theta_0)\right| \leq \sqrt{3(1+2\mu)\lambda_{I\!I}/N_J} \leq \sqrt{3(1+2\mu)\lambda_{I\!I}/(\rho N_{I\!I})}.$$

If the interval $I\!I$ is accepted, then there is a subinterval J of $I\!I$ such that (4.12) holds, and the assertion follows from the trivial inequality

$$\left|\ln(\widehat{\theta}/\theta_0)\right| \le \left|\ln(\widehat{\theta}/\tilde{\theta}_J)\right| + \left|\ln(\tilde{\theta}_J/\theta_0)\right|.$$

4.7.5 Proof of Theorem 4.7

In the homogeneous situation ($\theta_t = \theta$), our choice of the critical values λ_I ensures that with the probability al least $1-\beta_I$ it holds $L_J(\tilde{\theta}_J, \theta) = N_J K(\tilde{\theta}_J, \theta) \le \lambda_I/2$ for every interval $J \in \mathcal{J}(I)$, see the proof of Theorem 4.3. In the case of the change point model we get the same bound for all intervals of homogeneity $J \in \mathcal{J}(I)$ that does contain a change point. Below in the proof we now consider the situation with $N_J K(\tilde{\theta}_J, \theta) \le \lambda_I/2$ for all such intervals.

Let now $I\!I = [\nu, n[$ and $J = I \backslash I\!I$, so that $\theta_t = \theta$ for $t \in J$ and $\theta_t = \theta'$ for $t \in I\!I$. We therefore assume that $K(\tilde{\theta}_J, \theta) \le \lambda_I/(2m^*)$ and $K(\tilde{\theta}_{I\!I}, \theta') \le \lambda_I/(2m^*)$.

Denote $\theta_I = (\theta + \theta')/2$. Since $N_J = N_{I\!I} = m^*$ and $N_I = 2m^*$, it also holds that

$$\tilde{\theta}_I = (\tilde{\theta}_{I\!I} + \tilde{\theta}_J)/2.$$

The test statistic $T_{I,I\!I}$ can be represented as

$$T_{I,I\!I} = N_J K(\tilde{\theta}_J, \tilde{\theta}_I) + N_{I\!I} K(\tilde{\theta}_{I\!I}, \tilde{\theta}_I) = m^* d^2(\tilde{\theta}_J, \tilde{\theta}_{I\!I}).$$

Lemma 4.6 from below and the theorem condition imply that

$$d(\tilde{\theta}_J, \tilde{\theta}_{I\!I}) \ge d(\theta, \theta') - C_1\sqrt{\lambda_I/m^*} \ge \sqrt{\lambda_I/m^*}$$

and hence $T_{I,I\!I} \ge \lambda_I$. This completes the proof of the theorem.

4.7.6 Proof of Theorem 4.8

Let $u = \theta/\theta'$ and let $\tau \in \mathcal{T}_I$, $\tau < \nu$. Denote $m = \nu - \tau$, $D = [\tau, \nu[$ and $J = [\tau, n[$, $J^* = [\nu, n[$, $A = I \backslash J$ and $A^* = I \backslash J^*$. We aim to bound the

probability of the event $T_{I,\nu} < T_{I,\tau}$. More precisely, we intend to show that if m is sufficiently large then this probability is negligible. This particularly implies that the error of estimating the location of change point is bounded with a high probability.

Similarly to the proof of Theorem 4.7 we assume that $L_J(\tilde\theta_J, \theta) = N_J K(\tilde\theta_J, \theta) \le \lambda_{I'}/2$ for every $I' \in \mathcal{I}(I)$ and every $J \in \mathcal{J}(I')$ that does contain a change point. The probability of this event is not less than $1 - \alpha_I$. Obviously

$$\begin{aligned} T_{I,\nu} - T_{I,\tau} &= \widehat{L}_{A^*} + \widehat{L}_{J^*} - \left(\widehat{L}_A + \widehat{L}_J\right) \\ &= \widehat{L}_{J^*} + \widehat{L}_D - \widehat{L}_J + \widehat{L}_{A^*} - \widehat{L}_A - \widehat{L}_D \,. \end{aligned}$$

Since θ_t is constant for $t \in J^*$, it holds

$$\widehat{L}_{J^*} + \widehat{L}_D - \widehat{L}_J \le \lambda_I \,.$$

Next we show that $\widehat{L}_{J^*} + \widehat{L}_D - \widehat{L}_J > \lambda_I$. This would imply that $T_{I,\tau} < T_{I,\mathrm{cp}}$ and hence $|\widehat\nu - \nu| > m$ is impossible. Denote $\gamma = m/(m + N^*)$ where $N^* = N_{J^*} = n - \nu$. Without loss of generality we assume that $\gamma \le 1/2$. Define $d^2_\gamma(\theta, \theta') = \gamma K(\theta, \theta_\gamma) + (1-\gamma) K(\theta', \theta_\gamma)$ with $\theta_\gamma = \gamma\theta + (1-\gamma)\theta'$. Similarly to Lemma 4.6, it holds $\widehat{L}_{J^*} + \widehat{L}_D - \widehat{L}_J = N_J d^2_\gamma(\tilde\theta_{J^*}, \tilde\theta_D)$ and

$$d_\gamma(\tilde\theta_{J^*}, \tilde\theta_D) \ge d_\gamma(\theta_{J^*}, \theta_D) - C\sqrt{\lambda_I/m} \ge K^{1/2}(\theta, (\theta+\theta')/2) - C\sqrt{\lambda_I/m}$$

for some fixed C. Now the assertion easily follows from the conditions of the theorem.

4.7.7 Multivariate case

Here we give a sketch of the proof of Theorem 4.9. We consider a homogeneous multiple volatility case, when the vectors of returns R_t are i.i.d. from the multivariate normal distribution $\mathrm{N}(0, \Sigma)$ for all $t \in I = [n-m, n[$. The aim is to bound the corresponding test statistics $T_{I,\tau}$ for different internal points τ and also their supremum. Similarly to the univariate case, the key step is an extension of Theorem 4.10. With this extension, Theorem 4.9 can be proved in the same line as Theorem 4.1.

Theorem 4.12 *Let* $\Sigma_t = \Sigma_0$ *for all* $t \in I$ *where* Σ_0 *is a constant or a random variable measurable w.r.t.* $\mathcal{F}_{n-m-1}$. *Then it holds for any* Σ *and any* $z \geq 0$

$$\mathsf{P}\left\{L_I(\Sigma, \Sigma_0) \geq z\right\} \leq e^{-z}$$

and moreover,

$$\mathsf{P}\left\{L_I(\tilde{\Sigma}_I, \Sigma_0) \geq z + 1\right\} \leq N_I^{C_d} e^{-z}.$$

Proof The first assertion is very general in nature and it follows directly from the property of the log-likelihood: $\mathsf{E}_0 \exp L_I(\Sigma, \Sigma_0) = 1$ where E_0 means the measure corresponding to the volatility matrix Σ_0.

The proof of the second assertion is much more complicated. In particular, Lemma 4.1 which was essentially used in the proof of Theorem 4.1 does not hold in such a simple form. Below we establish an extension of this lemma for the multivariate case.

Without loss of generality, we assume that $\Sigma_0 = I_d$, that is, Σ_0 is the unit matrix and the returns are standard normal under the null hypothesis. Let $\mathcal{S}$ be the set of symmetric positive $d \times d$ matrices Σ. For two matrices Σ and $\bar{\Sigma}$ from $\mathcal{S}$, denote $l(\Sigma, \bar{\Sigma}) = -0.5\left\{\ln(\mathsf{det}\Sigma) + \mathsf{tr}(\Sigma^{-1} - I)\bar{\Sigma}\right\}$. Then $L_I(\Sigma, \Sigma_0) = N_I l(\Sigma, \tilde{\Sigma}_I)$.

Lemma 4.3 *For every* $z > 0$, *there exists a compact subset* $\mathcal{S}_z$ *of the set* $\mathcal{S}$ *such that*

$$\left\{\bar{\Sigma} : \sup_{\Sigma \in \mathcal{S}} l(\Sigma, \bar{\Sigma}) \geq z\right\} = \left\{\bar{\Sigma} : \sup_{\Sigma \in \mathcal{S}_z} l(\Sigma, \bar{\Sigma}) \geq z\right\}.$$

Proof let $\mathcal{U}$ be the set of all orthogonal $d \times d$ matrices. This set is obviously compact. Let also θ_* and θ^* be the values defined in Lemma 4.1 for the univariate case. Define $\mathcal{L}_z$ as a set of diagonal matrices with entries $\theta_i \in [\theta_*, \theta^*]$ for all $i = 1, \ldots, d$, and $\mathcal{S}_z = \mathcal{L}_z \times \mathcal{U}$. We show that

$$\left\{\bar{\Sigma} : \sup_{\Sigma \in \mathcal{S}} l(\Sigma, \bar{\Sigma}) \geq z\right\} = \left\{\bar{\Sigma} : \sup_{\Sigma \in \mathcal{S}_z} l(\Sigma, \bar{\Sigma}) \geq z\right\}$$

Any matrix $\Sigma \in \mathcal{S}$ can be decomposed as $\Sigma = \Gamma^\top \Lambda \Gamma$ where Γ is a orthogonal matrix, that is, $\Gamma^\top \Gamma = I_d$, and Λ is a diagonal matrix. Next, for every pair $\Sigma, \bar{\Sigma} \in \mathcal{S}$ holds $l(\Sigma, \bar{\Sigma}) = l(\Gamma \Sigma \Gamma^\top, \Gamma \bar{\Sigma} \Gamma^\top)$, and hence, we can reduce the proof to the case when $\bar{\Sigma}$ runs over the set of diagonal matrices. For a fixed $\bar{\Sigma}$, the maximum of $l(\Sigma, \bar{\Sigma})$ over $\Sigma \in \mathcal{S}$ is attained at $\Sigma = \bar{\Sigma}$. Thus, if $\bar{\Sigma}$ are restricted to the set of diagonal matrices $\mathcal{L}$, the same can be assumed for Σ, and it suffices to show that

$$\left\{ \bar{\Lambda} \in \mathcal{L} : \sup_{\Lambda \in \mathcal{L}} l(\Lambda, \bar{\Lambda}) \geq z \right\} = \left\{ \bar{\Lambda} \in \mathcal{L} : \sup_{\Lambda \in \mathcal{L}_z} l(\Lambda, \bar{\Lambda}) \geq z \right\}.$$

It remains to note that for $\bar{\Lambda} = \mathsf{diag}(\bar{\lambda}_1, \ldots, \bar{\lambda}_d)$ and $\Lambda = \mathsf{diag}(\lambda_1, \ldots, \lambda_d)$

$$l(\Lambda, \bar{\Lambda}) = \sum_{i=1}^{d} l_1(\lambda_i, \bar{\lambda}_i)$$

with $l_1(\lambda, \bar{\lambda}) = -0.5 \left(\ln \lambda + (\lambda^{-1} - 1) \bar{\lambda} \right)$ and the result follows from the similar assertion for the univariate case in Lemma 4.1.

For every two matrices $\Sigma, \bar{\Sigma} \in \mathcal{S}_z$, it obviously holds $L_I(\Sigma, \Sigma_0) - L_I(\bar{\Sigma}, \Sigma_0) = 0.5 N_I \mathsf{tr} \left(\Sigma^{-1} \tilde{\Sigma}_I - (\bar{\Sigma})^{-1} \tilde{\Sigma}_I \right)$. Define $\rho^2(\Sigma, \bar{\Sigma}) = \mathsf{tr} \big(\Sigma^{-1} - (\bar{\Sigma})^{-1} \big)^2$. By the Cauchy-Schwarz inequality

$$|L_I(\Sigma, \Sigma_0) - L_I(\bar{\Sigma}, \Sigma_0)| \leq 0.5 N_I \rho(\Sigma, \bar{\Sigma}) \big(\mathsf{tr} \tilde{\Sigma}_I^2 \big)^{1/2}.$$

For a sufficiently large constant a, the value $\big(\mathsf{tr} \tilde{\Sigma}_I^2 \big)^{1/2}$ is bounded by $2 N_I^{a-1}$ with a high probability. This yields (up to a set of a vanishing probability) that for all $\Sigma, \bar{\Sigma}$ with $\rho(\Sigma, \bar{\Sigma}) \leq \epsilon \stackrel{\text{def}}{=} N_I^{-a}$ it holds $|L_I(\Sigma, \Sigma_0) - L_I(\bar{\Sigma}, \Sigma_0)| \leq 1$.

Let $\mathcal{D}_\epsilon$ be the ϵ-net in the compact set $\mathcal{S}_z$ w.r.t. the metric $\rho(\cdot, \cdot)$. It is easy to see that this set is finite with the cardinality of order $n^{d^2 a}$. Then

$$\begin{aligned} \mathsf{P} \left\{ L_I(\tilde{\Sigma}_I, \Sigma_0) \geq z + 1 \right\} &\leq \mathsf{P} \left\{ \sup_{\Sigma \in \mathcal{D}_\epsilon} L_I(\Sigma, \Sigma_0) \geq z \right\} \\ &\quad + \mathsf{P} \left\{ \sup_{\rho(\Sigma, \bar{\Sigma}) \leq \epsilon} |L_I(\Sigma, \Sigma_0) - L_I(\bar{\Sigma}, \Sigma_0)| > 1 \right\} \leq \sum_{\Sigma \in \mathcal{D}_\epsilon} e^{-z} = n^{d^2 a} e^{-z} \end{aligned}$$

as required.

4.7.8 Some lemmas

In this section we collect some technical facts about the properties of the normal family with varying variance.

Lemma 4.4 *Let R be normal with parameters $(0, \theta)$ for some $\theta > 0$. Then for any θ' such that $|\theta/\theta' - 1| \le 0.8 \min\{\mu, 1\}$ it holds*

$$\ln \mathsf{E} \exp \mu^{-1} \{\ell(R, \theta) - \ell(R, \theta')\} \le \mu^{-2} (\theta/\theta' - 1)^2$$

and

$$\begin{aligned} \ln \mathsf{E} \exp \mu^{-1} \{\ell(R, \theta) - \ell(R, \theta') - K(\theta, \theta')\} &\le \mu^{-2} (\theta/\theta' - 1)^2 \\ \ln \mathsf{E} \exp \mu^{-1} \{\ell(R, \theta') - \ell(R, \theta) + K(\theta, \theta')\} &\le \mu^{-2} (\theta/\theta' - 1)^2 \end{aligned}$$

where $K(\theta, \theta') = -0.5 \{\ln(\theta/\theta') - 1 + \theta/\theta'\}$.

Proof Denote $\delta = \theta/\theta' - 1$. Since $\xi = \theta^{-1/2} R$ is standard normal, it holds

$$\begin{aligned} \ln \mathsf{E} \exp \mu^{-1} \{\ell(R, \theta) - \ell(R, \theta')\} &= \frac{1}{2\mu} \ln(\theta'/\theta) + \ln \mathsf{E} \exp \left\{ R^2 (1/\theta' - 1/\theta)/(2\mu) \right\} \\ &= -\frac{1}{2\mu} \ln(\theta/\theta') + \ln \mathsf{E} \exp \left\{ \xi^2 \delta/(2\mu) \right\} \\ &= -\frac{1}{2\mu} \ln(1 + \delta) - \frac{1}{2} \ln(1 - \delta/\mu). \end{aligned}$$

Since $|\ln(1 - u) + \mu^{-1} \ln(1 + \mu u)| \le 2u^2$ for $|u| \le 0.8$ and $|\mu u| \le 0.8$, the first assertion of the lemma follows.

Similarly

$$\begin{aligned} \ln \mathsf{E} \exp \mu^{-1} \{\ell(R, \theta) - \ell(R, \theta') - K(\theta, \theta')\} &= -\frac{\delta}{2\mu} - \frac{1}{2} \ln(1 - \delta/\mu) \le \delta^2/\mu^2 \\ \ln \mathsf{E} \exp \mu^{-1} \{\ell(R, \theta') - \ell(R, \theta) + K(\theta, \theta')\} &= \frac{\delta}{2\mu} - \frac{1}{2} \ln(1 + \delta/\mu) \le \delta^2/\mu^2. \end{aligned}$$

Lemma 4.5 *The inequality $u - \ln(1 + u) \le 2\delta^2$ for some $\delta \ge 0$ implies for all $u \ge -1/2$ that $|\ln(1 + u)| \le \sqrt{6}\delta$.*

Proof Denote $x = \ln(1+u)$. Then, for $u \geq 0$, it holds $u - \ln(1+u) - 1/2 \ln^2(1+u) = e^x - 1 - x - x^2/2 \geq 0$, that is, $\ln^2(1+u) \leq 2u - 2\ln(1+u) \leq 4\delta^2$. For $u \in [-1/2, 0]$, one similarly gets $u - \ln(1+u) - (1/3)\ln^2(1+u) = e^x - 1 - x - x^2/3 \geq 0$.

The next result concerns the distance $d(u,v)$ introduced in Section 4.5.6: $d^2(u,v) = K(u,w) + K(v,w)$ for $w = (u+v)/2$.

Lemma 4.6 *There exists a constant C_1 such that for any positive numbers u, u_0, v, v_0, and for any $\delta \in [0,1]$, the conditions $K(u,u_0) \leq \delta^2/2$, $K(v,v_0) \leq \delta^2/2$ imply*

$$d(u,v) \geq d(u_0,v_0) - C_1\delta.$$

Proof It sufficient to check that the functions $d(u,v)$, $K^{1/2}(u,u_0)$ and $K^{1/2}(v,v_0)$ have bounded partial derivatives w.r.t. the both variables u, v. We omit the details.

Bibliography

Andersen, T. G., Benzoni, L., and Lund, J. (2002). An empirical investigation of continuous-time equity return models. *Journal of Finance*, 3:1239–1284.

Baillie, R., Bollerslev, T., and Mikkelsen, H. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74:3–30.

BIS (1996). Supervisory framework for the use of "backtesting"in conjunction with the internal models approach to market risk capital requirements. Basel Commitee on Banking Supervision.

Bollerslev, T. (1986). Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, S. Francisco.

Brockmann, M., Gasser, T., and Herrmann, E. (1993). Locally adaptive bandwidth choice for kernel regression estimators. *Journal of the American Statistical Association*, 88:1302–1309.

Brodskij, B. and Darkhovskij, B. (1993). *Nonparametric methods in change point problems*. Mathematics and its Applications. Kluwer Academic Publisher, Dordrecht.

Cai, Z., Fan, J., and Li, R. (2000a). Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, 95:888–902.

Cai, Z., Fan, J., and Yao, Q. (2000b). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95(451):941–956.

Canarella, G., Pollard, S. K., and Lai, K. S. (1990). Cointegration between exchange rates and relative prices: Another view. *European Economic Review*, 34:1303–1322.

Carroll, R. and Ruppert, D. (1988). *Transformation and Weighting in Regression*. Chapman and Hall, New York. Chapter 4.

Cheng, M. Y., Fan, J., and Spokoiny, V. (2003). Dynamic nonparametric filtering with application to volatility estimation. In Acritas and Politis, editors, *Recent Advances and Trends in Nonparametric Statistics*, pages 315–333. Elsevier.

Christoffersen, P. and Giorgianni, L. (2000). Interest rate in currency basket: Forecasting weights and measuring risk. *Journal of Business and Economic Statistics*, 18:321–335.

Clements, M. P. and Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge University Press, Cambridge.

Clements, M. P. and Hendry, D. F. (1999). *Forecasting Non-stationary Economic Time Series*. MIT Press, Cambridge.

Cooley, T. F. and Prescott, E. C. (1973). An adaptive regression model. *International Economic Review*, 14:364–371.

Dahlhaus, R. and Rao, S. S. (2003). Statistical inference of time-varying ARCH processes. DFG Schwerpunktprogramm 1214, Universität Bremen, Preprint Nr. 31.

Deutsch, H. P. (2001). *Derivate und Interne Modelle*. Schäffer-Poeschel Verlag, Stuttgart, 2 edition.

Donoho, D., Johnstone, I., Kerkyacharian, G., and Picard, D. (1994). Wavelet shrinkage: asymptopia ? *Journal of the Royal statistical Society*, 57:301–369. Series B.

Duffie, D., Pan, J., and Singleton, K. (2000). Transform analysis and asset pricing for affine jump diffusions. *Econometrica*, 68:1343–1376.

Eberlein, E. and Prause, K. (2002). The generalized hyperbolic model: financial derivatives and risk measures. In Geman, H., Madan, D., Pliska, S., and Vorst, T., editors, *Mathematical Finance-Bachelier Congress 2000*, pages 245–267, Berlin. Springer Verlag.

Eichengreen, B., Masson, P., Savastano, M., and Sharma, S. (1999). *Transition Strategies and Nominal Anchors on the Road to Greater Exchange Rate Flexibility*. Princeton University Press, Princeton.

Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, 50:987–1008.

Engle, R. F., editor (1995). *ARCH, selected readings*. Oxford University Press, Oxford.

Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometrc Reviews*, 5:1–50, 81–87.

Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society*, 57:371–394.

Fan, J. and Gu, J. (2003). Semiparametric estimation of Value at Risk. *Econometrics Journal*, 6:260–289.

Fan, J., Jiang, J., Zhang, C., and Zhou, Z. (2001). Time-dependent diffusion models for term structure dynamics and the stock price volatility. *Statistica Sinica*, 27.

Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficent models. *Annals of Statistics*, 27:1491–1518.

Franke, J., Härdle, W., and Hafner, C. (2003). *Einführung in die Statistik der Finanzmärkte*. Springer Verlag, Berlin, second edition.

Franses, P. and Dijk, D. V. (1996). Forecasting stock market volatility using (non-linear) GARCH models. *Journal of Forecasting*, 15:229–235.

Glosten, L., Jagannathan, R., and Runkle, D. (1992). On the relation between the expected value and the volatility of nominal excess return on on stocks. *Journal of Finance*, 46:1779–1801.

Gouriéroux, C. (1997). *ARCH Models and Financial Application*. Springer Verlag, Berlin.

Granger, C. W. J. (1986). Developments in the study of cointegrated economic variables. *Oxford Bulletin of Economics and Statistics*, 48:213–228.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalised Linear Models*, volume 58 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357–384.

Hamilton, J. D. (1993). *Handbook of Statistics*, volume 9, chapter 3: "Estimation, Inference, and Forecasting of Time Series Subject to Changes in Regime". Elsevier.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton.

Hamilton, J. D. and Baldev, R., editors (2002). *Advances in Markov-Switching Models*. Springer Verlag, Berlin.

Hansen, B. E. (1992). Tests for parameter instability in regressions with I(1) processes. *Journal of Business and Economic Statistics*, 10:242–254.

Härdle, W., Herwartz, H., and Spokoiny, V. (2003). Time inhomogeous multiple volatility modelling. *Journal of Financial Econometrics*, 1:55–95.

Härdle, W., Spokoiny, V., and Teyssière, G. (2000). Adaptive estimation for a time inhomogeneouse stochastic volatility model. Discussion Paper 6, Sonderforschungsbereich 373, Humboldt-Universität zu Berlin.

Härdle, W. and Stahl, G. (2000). Backtesting beyond VaR. In Franke, J., Haerdle, W., and Stahl, G., editors, *Measuring Risk in Complex Stochastic Systems*, Lecture Notes in Statistics. Springer Verlag.

Harvey, A., Ruiz, E., and Shephard, N. (1995). Multivariate stochastic variance models. In Engle (1995).

Härdle, W., Müller, M., Sperlich, S., and Werwatz, A. (2004). *Nonparametric and Semiparametric Models*. Springer Verlag, Berlin.

Kitagawa, G. (1987). Non-Gaussian state-space modelling of nonstationary time series. *Journal of the American Statistical Association*, 82:1032–1063.

Kleibergen, F. and Dijk, H. K. V. (1993). Nonstationarity in garch models: A bayesian analysis. *Journal of Applied Econometrics*, 8:41–61.

Lamureux, C. G. and Lastrapes, W. D. (1990). Persistence in variance, structural change and the garch model. *Journal of Business and Economic Statistics*, 8:225–234.

Lepski, O. (1990). One problem of adaptive estimation in gaussian white noise. *Theory of Probability and Applications*, 35:459–470.

Lepski, O. and Spokoiny, V. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Annals of Statistics*, 25:2512–2546.

Lipster, R. and Spokoiny, V. (1999). Deviation probability bound for martingales with applications to statistical estimation. *Statistics & Probability Letters*, 46:347–357.

Lütkepohl, H. (1988). Prediction tests for structural instability. *Journal of Econometrics*, 39:267–296.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroskedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7:271–300.

Meese, R. and Rogoff, K. (1983). Empirical exchange rate models of the seventies: Do they fit out of the sample ? *Journal of International Economics*, 14:3–24.

Mercurio, D. and Spokoiny, V. (2004). Statistical inference for time-inhomogeneous volatility models. *Annals of Statistics*, 32:577–602.

Mercurio, D. and Torricelli, C. (2003). Estimation and arbitrage opportunities for exchange rate baskets. *Applied Econometrics*, 35:1689–1698.

Mikosch, T. and Starica, C. (2000a). Change of structure in financial time series, long range dependence and the garch model. University of Aarhus, Aarhus school of Business, Discussion Paper Nr. 58.

Mikosch, T. and Starica, C. (2000b). Is it really long memory we see in financial returns? In Embrechts, P., editor, *Extremes and Integrated Risk Management*, Risk Books. UBS Warburg.

Mussa, M., Masson, P., Swoboda, A., Jadresic, E., Mauro, P., and Berg, A. (2000). Excahnge rate regimes in an increasingly integrated world economy. International Monetary Fund, Occasional Paper Nr. 193.

Nelson, D. B. (1995). Conditional heteroscedasticity in asset returns: A new approach. In Engle (1995).

Pollak, M. (1985). Optimal detection of a change in distribution. *Annals of Statistics*, 13:206–227.

Schäl, M. (1994). On quadratic cost criteria for option hedging. *Mathematics of Operation Research*, 19:121–133.

Sentana, E. (1995). Quadratic ARCH models. *Review of Economic Studies*, 62:639–661.

Simonato, J. G. (1992). Estimation of GARCH processes in the presence of structural change. *Economics Letters*, 40:155–158.

Spokoiny, V. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *Annals of Statistics*, 26:1356–1378.

Starica, C. (2003). Is GARCH(1,1) as good a model as the Nobel prize accolades would imply? unpublished manuscript, `www.math.chalmers.se/ starica/papersoft.pdf`.

Stock, J. H. (1993). *Handbook of Econometrics*, volume 4, chapter 46: "Unit Roots, Structural Breaks and Trends". Elsevier.

Taylor, S. J. (1986). *Modelling Financial Time Series*. Wiley, Chichester.

Wang, Y. (1995). Jump and sharp cusp detection by wavelets. *Biometrica*, 82:385–397.

Lebenslauf

Name:	Danilo Mercurio
10/1996 - 07/1999	Studium an der Humboldt-Universität zu Berlin in der Fachrichtung Wirtschaftswissenschaften
10/1999 - 07/2004	Promotionsstudium an der Humboldt-Universität zu Berlin in der Fachrichtung Wirtschaftswissenschaften, Arbeitsgebiet: Statistik der Finanzmärkte
09/2002 - 05/2003	Masterstudium an der Columbia University - New York in der Fachrichtung Finanzmathematik
07/2003 - 12/2003	Wissenschaftlicher Mitarbeiter an der Humboldt-Universität zu Berlin, Sonderforschungsbereich 373
08/2003 - 07/2004	Wissenschaftlicher Mitarbeiter an dem Weierstraß Institut in der Forschungsgruppe Stochastische Algorithmen und Nichtparametrische Statistik

Selbständigkeitserklärung

Bei der Erstellung meiner Dissertation habe ich ausser der angegebenen Literatur keine weitere Literatur benutzt. Weiterhin ist mir bei der Erstellung dieser Arbeit ausschliesslich von den Personen Hilfe zuteil geworden, die in der Danksagung genannt werden.
Ich bezeuge durch meine Unterschrift, dass meine Angaben über die bei der Abfassung meiner Dissertation benutzten Hilfsmittel, über die mir zuteil gewordene Hilfe sowie über frühere Begutachtungen meiner Dissertation in jeder Hinsicht der Wahrheit entsprechen.

Danilo Mercurio
10. Mai 2004